Firm Growth through New Establishments*

Dan Cao[†] Georgetown University

Henry R. Hyatt[‡] U.S. Census Bureau

Toshihiko Mukoyama[§] Georgetown University

Erick Sager[¶] Federal Reserve Board

December 2019

Abstract

This paper analyzes the distribution and growth of firm-level employment along two margins: the extensive margin (the number of establishments in a firm) and the intensive margin (the number of workers per establishment in a firm). We utilize administrative datasets to document the behavior of these two margins in relation to changes in the U.S. firm-size distribution. In the cross section, we find the firm-size distribution, as well as both extensive and intensive margins, exhibits a fat tail. The increase in average firm size between 1990 and 2014 is primarily driven by an expansion along the extensive margin, particularly in very large firms. We develop a tractable general equilibrium growth model with two types of innovations: external and internal. External innovation leads to the extensive margin of firm growth, and internal innovation leads to intensive-margin growth. The model generates fat-tailed distributions in firm size, establishment size, and the number of establishments per firm. We estimate the model to uncover the fundamental forces that caused the distributional changes from 1995 to 2014. We find the change in the external innovation cost and the decline in establishment exit rates are the largest contributors to the increase in the number of establishments per firm.

Keywords: firm growth, firm-size distribution, establishment, innovation

JEL Classifications: E24, J21, L11, O31

---

*First version: February 2017. We thank Jess Helfand and Mike LoBue for their help with Quarterly Census of Employment and Wages data. We also thank helpful comments and suggestions from Ufuk Akcigit, Sina Ates, Andy Atkeson, Nick Bloom, Stephane Bonhomme, John Earle, Jeremy Greenwood, Hubert Janicki, Sam Kortum, Makoto Nirei, David Ratner, John Shea, and Mike Siemer, as well as the seminar participants at Arizona State University, Federal Reserve Bank of Atlanta, George Mason University, Hitotsubashi University, IMF, Loughborough University, North Carolina State University, Queensland University, University of Alberta, University of California at San Diego, University of Connecticut, University of Illinois, University of Maryland, University of Texas at Austin, University of Tokyo, University of Warwick, Uppsala University, and conference participants at Barcelona GSE 2018, CICM conference 2018 in Beijing, GCER Biennial Alumni Conference 2019, Midwest Macro 2018 at the University of Wisconsin, SAET 2018 in Taiwan, SAET 2019 in Ischia, SCE 2018 in Milan, and SED 2019 in St Louis. We thank Bobbie Goettler for editorial help. This research was conducted with restricted access to microdata from the U.S. Bureau of Labor Statistics and U.S. Census Bureau. All results have been reviewed to ensure no confidential data are disclosed. Results using U.S. Census Bureau microdata have been approved by the Disclosure Review Board with authorization number DRB-B0063-CED-20190628. The views expressed within are those of the authors and do not necessarily reflect those of the Federal Reserve System, the Census Bureau, the Bureau of Labor Statistics or the U.S. government. All errors are our own.

[†]Contact information: dan.cao@georgetown.edu

[‡]Contact information: henry.r.hyatt@gmail.com

[§]Contact information: tm1309@georgetown.edu

[¶]Contact information: erick.sager@gmail.com

1 Introduction

Understanding the process of firm growth is essential in the analysis of macroeconomic performance. Firms that innovate and expand are the driving force of output and productivity growth. Recent analysis of macroeconomic productivity emphasizes the role of innovation and reallocation at the firm level, from both the theoretical and empirical standpoint.

In this paper, we focus on a particular aspect of firm growth: growth through adding new establishments. A firm can increase its size along two margins; it can add more workers to its existing establishments or it can build new establishments. We call the former the *intensive margin* of firm growth and the latter the *extensive margin* of firm growth. This distinction is important because these margins typically imply different reasons for expansion. A new manufacturing plant is often built to produce a new product. In the service sector, building a new store or a new restaurant implies venturing into a new market. Creating a new establishment is also different in that it typically requires a significant amount of investment in equipment and structures. In the following, we also utilize the terminology of extensive and intensive margins in describing the firm-size distribution.

Given the importance of this distinction, the fact that very little is known about how firms grow through new establishments is somewhat surprising. Our first goal is to establish stylized facts about firm growth in these two margins. We describe the firm-size distribution along these margins, and document how these two margins have changed in recent years. We then interpret the facts through the lens of a macroeconomic model of endogenous firm dynamics.

In the existing literature on firm dynamics, firms and establishments are frequently treated as interchangeable. One justification for this assumption is the fact that the majority of firms are single-establishment firms. Although this assumption is justifiable in some situations, it is misleading in many macroeconomic contexts. In the U.S. economy, although 95% of firms are single-establishment firms, their share in total employment is less than half (45%).[1] Furthermore, the firm-size distribution exhibits a Pareto tail, which implies large firms with many establishments have a disproportionately large impact on macroeconomic performance (Gabaix, 2011). Therefore, understanding how these large firms are created is an important research question.

As in the literature, we find that the firm-size distribution has a Pareto tail in our dataset. Our novel finding is that the two margins of the firm-size distribution, the extensive margin (the number of establishments per firm) and the intensive margin (the average number of workers per establishment), also exhibit Pareto tails.

Along the time series, we find the average firm size has grown in recent years. This trend is largely driven by firms in the right tail of the size distribution. This observation is consistent with the findings of recent studies.[2] These studies suggest this increase in size has had important

[1] The numbers are taken from 2005 Q1 Business Employment Dynamics (which is drawn from the Quarterly Census of Employment and Wages, explained in Section 2.1): `https://www.bls.gov/bdm/sizeclassqanda.htm`. In the manufacturing sector, single-plant firms own 72% of plants, they produce only 22% of the value added (Kehrig and Vincent, 2019).

[2] For example, Autor et al. (2017) document the emergence of superstar firms. De Loecker and Eeckhout (2017)

implications on other changes in macroeconomic variables, such as the decline in labor share. Our new finding compared to the literature is that this expansion is driven by the extensive margin.

To investigate what changes in the economic environment have contributed to this phenomenon, we build a macroeconomic model of endogenous firm growth. Our model extends previous work by Klette and Kortum (2004) and Luttmer (2011). In these papers, each individual firm grows by adding production units ("product lines" in Klette and Kortum (2004) and "blueprints" in Luttmer (2011)). In our model, we call this margin of firm growth *external innovation*. These production units can be naturally interpreted as establishments. Thus, this type of framework provides an ideal vehicle for analyzing firm growth through new establishments. The major departure of our model, compared to these two papers, is the recognition that each establishment can grow in size.[3] We introduce this technological improvement at the establishment level, which we call *internal innovation*, and explicitly compare our model outcomes with the data on establishments.

We estimate model parameters to match U.S. firm-size distributions in 1995 and 2014. Through the lens of our model, we uncover fundamental forces that have caused the changes in the firm size distribution and its components over this time period. The estimation shows (i) the external innovation cost became lower for firms that are active in such innovations, (ii) the internal innovation cost became higher, (iii) entry became more costly, (iv) the average quality at entry became higher, and (v) upon entry, growth through external innovation became more common but shorter-lived. We also find the largest contributors to the increase in the number of establishments per firm are the change in the external innovation cost and the decline in the establishment exit rate. The estimated model allows us to conduct counterfactual experiments. These experiments reveal the quantitative importance of each parametric change in generating the observed patterns in the data.

Akcigit and Kerr's (2018) model has features similar to the model in this paper. They also consider innovations that are internal to the establishments ("products" in their terminology) that the firm already operates and external innovations that increase the number of establishments that the firm operates, although many of their model assumptions are different from ours.[4] Our main purpose is to map the model to the data on firm and establishment sizes, measured by employment, whereas Akcigit and Kerr (2018) primarily use patent data and therefore consider products rather than establishments.[5] Our model is tractable and it allows for analytical characterizations of Pareto tails in the distributions of firm size, establishment size, and number of establishments per firm. By contrast, Akcigit and Kerr's (2018) model does not allow for Pareto tails. Our model is customized

analyze the increase in concentration measured by the markup. Gutiérrez and Philippon (2017) also point out the recent increase in concentration in the U.S. corporate sector.

[3]Acemoglu and Cao (2015) have also pointed out such growth is lacking in the Klette and Kortum (2004) framework.

[4]For example, in Akcigit and Kerr (2018) an external innovation improves a product that is already produced by another firm, whereas in our model, it creates a new establishment whose quality is the same as the other establishments in the firm. Arguably, our assumption better suits the service-sector innovations. When a retail firm opens a new store after researching the placement of a new location, the quality of the new store would more likely be associated with the firm that opens the store than the existing stores in that location.

[5]In the context of innovation-based growth models, Garca-Macia et al. (2019) and Mukoyama and Osotimehin (2019) employ similar estimation strategies as ours in different settings

to our question—we investigate the forces behind the changes in firm size distribution in the recent years. In particular, we highlight the role of changes in innovation costs in allowing firms to expand on the extensive margin.

Lentz and Mortensen (2008) extends Klette and Kortum's (2004) model and estimates it using Danish data. Whereas their main focus is on productivity and reallocation, this paper mainly targets labor market facts. We directly exploit establishment-level information in analyzing firm dynamics. Furthermore, the empirical phenomenon we highlight mainly concerns large firms, and our model is tailored to fit the right tail of the firm-size distribution. Lentz and Mortensen (2008) match many salient features of the firm-size distribution, but misses the heaviness of the right tail in the data; the present paper fills this gap.[6]

Consistent with our estimated result, Aghion et al. (2019) argue the cost of firm expansion (a flow overhead cost in their model) has declined in recent years. Their focus is on the trend of productivity growth and the labor share. By contrast, our focus is on the change in the pattern of firm growth over time and analyzing the tail properties of the firm-size distribution.[7] Hsieh and Rossi-Hansberg (2019) show the industry-level increase in concentration is associated with the extensive-margin growth of top firms. They argue that cause of the extensive-margin growth of top firms in the service industry is driven by a new technology with a large fixed costs, such as Information and Communications Technology.

Finally, our paper is related to the theoretical literature studying the firm-size distribution and its fat right tail.[8] Luttmer (2011) and Acemoglu and Cao (2015) build endogenous growth models with fat (Pareto) tail distributions arising from only extensive-margin growth (firm adding establishments in the former) or intensive-margin growth (establishments expanding their size in the latter). We build a model with both margins to match their characteristics in the data. We are able to characterize the Pareto-tail indexes analytically, employing recent advances in the literature on regular variations and their applications.[9]

The paper is organized as follows. Section 2 describes the empirical patterns of firm growth in our dataset. Section 3 sets up the model, and Section 4 provides a theoretical characterization. Section 5 provides an additional characterization of the model's stationary firm-size distributions, including extensive- and intensive-margin distributions. Section 6 estimates model parameters and uses the model to perform a quantitative decomposition of 1990-2015 U.S. firm growth. Section 7 concludes.

[6] Lentz and Mortensen (2008) note that "While generally performing well in terms of matching firm size distributions the problem fitting the heavy far right tail in empirical size distribution is a well-known issue associated with the Klette and Kortum (2004) model. Improvement of the model along this dimension is a topic well worthy of future research" (pp. 1346-1347).

[7] Akicigit and Ates (2019) emphasizes the role of knowledge diffusion in the recent rise of concentration.

[8] See, for example, Axtell (2001), Gabaix (2009), Axtell and Guerrero (2019), and Kondo et al. (2019) for empirical studies.

[9] See, for example, Bingham et al. (1987), Mimica (2016), Gabaix et al. (2016), and Beare and Toda (2017).

2 Empirical facts

We first describe the data sources. After clarifying the conceptual framework, we present cross-sectional and time series facts related to firm size, establishment size at each firm, and number of establishments per firm.

2.1 Data

We use establishment-level microdata for the U.S. that report payroll information on the number of workers by establishment and month. Each establishment in the data has an employer identification number (EIN) and a 6-digit NAICS code associated with its primary industry.

This paper utilizes two restricted access datasets that use common source data that contains a near census of establishments in the U.S. The source data are collected for the Quarterly Census of Employment and Wages (QCEW) by U.S. states in partnership with the Bureau of Labor Statistics (BLS) for the official administration of state unemployment insurance programs. States also provide establishment-level QCEW microdata to the U.S. Census Bureau's Longitudinal Employer-Household Dynamics program as part of the Local Employment Dynamics federal-state partnership.[10] For each figure and table, we explicitly indicate the data source that is used.

We use the employer identification number (EIN) as the definition of the firm. This level is the one at which companies file their tax returns, and recent studies often consider it as the boundary of the firm. In some cases, EINs can be different from the ultimate ownership, especially for large firms. Song et al. (2015) discuss this issue at length in their study of inequality. They state that for 4,233 New York Stock Exchange publicly listed firms in the Dunn & Bradstreet database, 13,377 EINs are reported. They provide some examples: Walmart operates with separate EINs for "Walmart Stores," Supercenters, Neighborhood Markets, Sam's Clubs, and On-line divisions. Stanford University has separate EINs for the university, the bookstore, the main hospital, and children's hospitals. We view the EIN as a reasonable definition of a firm for our analysis, given that it is an economically meaningful unit (especially from the accounting perspective), and given that typically the EINs for large firms with multiple EINs are still substantially more aggregated than establishments.[11]

2.2 Conceptual framework for the firm-size decomposition

We define *firm size* as the total number of workers employed by a firm. In what follows, we decompose average firm size into an *extensive and intensive margin*. The *extensive margin* is the

[10] Data from the BLS contains information on 38 states, and data from the U.S. Census Bureau contains information on 28 states. Data from the U.S. Census Bureau are contained in the Employer Characteristics File maintained by the Census Bureau's LEHD program. See Abowd et al. (2009) and Vilhuber (2018), and Appendix D for further details of these datasets. In using data from both BLS and Census sources, we have confirmed using basic statistics that no inconsistencies exist across the two data frames.

[11] For example, as of October 31 2017, Walmart has 402 Discount Stores, 3,552 Supercenters, 702 Neighborhood Markets, 660 Sam's Clubs, and 97 Small Formats including E-Commerce Acquisition/C-stores (numbers are taken from `https://corporate.walmart.com/our-story/locations/united-states`).

total number of establishments owned by a firm, and the *intensive margin* is the average size of establishments owned by a firm (e.g., the number of workers per establishment in a firm).

For the sake of exposition, suppose F_t firms exist at time t, indexed by i. Let N_{it} be the total number of workers employed by firm i and let E_{it} be the total number of establishments owned and operated by firm i at time t. We define average establishment size within firm i as N_{it}/E_{it} so that we decompose size at the firm-level as

$$N_{it} = \left(\frac{N_{it}}{E_{it}}\right) \times E_{it}.$$

Accordingly, our measures of average firm size, the average intensive margin and the average extensive margin are

$$\left(\frac{1}{F_t}\sum_{i=1}^{F_t} N_{it}, \frac{1}{F_t}\sum_{i=1}^{F_t} \frac{N_{it}}{E_{it}}, \frac{1}{F_t}\sum_{i=1}^{F_t} E_{it}\right),$$

respectively.

Publicly-available datasets such as Business Dynamics Statistics of the U.S. Census Bureau contain the size distributions of establishments and firms separately. Compared to studies that utilize these publicly-available datasets, our study has several advantages. First, the use of microdata enables us to characterize the entire size distribution of firms, particularly at the right tail. Second, we cannot decompose firm growth into intensive and extensive margins without the information contained in the microdata we utilize. Whereas the average size and average extensive margin can be computed from the publicly-available information (the former is total employment divided by the total number of firms, and the latter is the total number of establishments divided by the total number of firms), the intensive margin cannot be computed from the publicly-available information.[12]

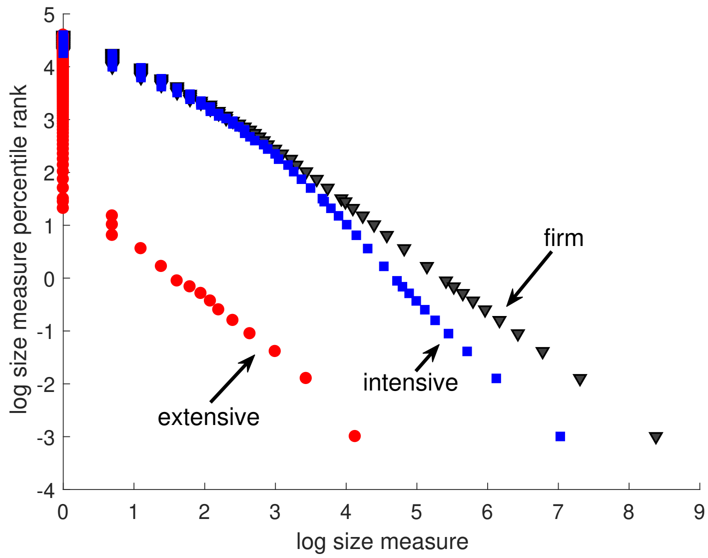
2.3 Cross-sectional properties

We first describe the cross-sectional distributions of firm size, the intensive margin, and the extensive margin. For this analysis, we use LEHD microdata and focus on 2005. Figure 1 plots the complementary cumulative distribution function in log-log scale, a type of figure commonly used in the literature to demonstrate whether the data are consistent with Pareto's Law.[13] All three series have a right tail that can be approximated by a straight line. This fact implies all three distributions have Pareto tails. To our knowledge, out paper is the first in the literature to document that both the extensive margin and intensive margin have Pareto tails. This fact is important for our model analysis later, because it indicates a similar mechanism is at work in intensive-margin

[12]Previous research has used publicly available data to compute an approximation to the intensive margin as the ratio of the total number of workers to the total number of establishments. However, this approximation fails for larger firms because it incorporates a correlation between size and establishment number that should not be contained in the decomposition. See Appendix E.1 for a formal treatment.

[13]Figure 1 presents rounded predicted values constructed from polynomial approximations of the size-rank relationship present in LEHD microdata. For the set of polynomial estimates, see Appendix Table D.2.1.

Figure 1: Size-rank relationships, ranked separately by size measure



Notes:Author's calculations of LEHD microdata. Each series shows predicted the predicted value of employment from a regression of the log size measure on a fifth order polynomial of the log percentile rank of the size measure in March of 2005, rounded to the nearest integer. See text for additional details.

growth and extensive-margin growth. In particular, the existence of Pareto tails indicates Gibrat's Law holds for large sizes, and this imposes some structure on the driving forces behind the growth in both margins.[14]

Note that the intensive-margin distribution is conceptually different from the economy-wide distribution of establishment size, although they are closely related. The intensive margin is a *firm-level concept*. When a firm with 100 establishments exists, for example, these establishments would count as 100 observations in the establishment-size distribution, whereas it is one sample point in our description of the intensive margin. Our empirical description here focuses on the intensive margin given that our interest in this paper is in firm size and firm growth. Although some existing studies have already found the U.S. establishment-size distribution exhibits a Pareto tail, this study is the first to describe the right-tail properties of the intensive margin of the firm-size distribution.

The existence of Pareto tails indicates firms at the top of the size distribution employ a greater share of all workers and own a greater share of establishments relative to the rest of the distribution.[15] Accordingly, the data points in Figure 1 imply 55.1% of employment is concentrated in the

[14] Figure 1 is also useful for understanding the entire distributions of the firm size, the extensive margin, and the intensive margin. Note that the vertical axis shows the log of the percentile rank, which ranges from 0 to 100. The value of log(100) is 4.6, and this is the upper bound of the vertical axis. Because $\log(1) = 0$, the values below zero indicate the sizes above the 99 percentile. In particular, because $\exp(-3) = 0.05$, the points with the vertical axis being -3 correspond to 99.95 percentile.

[15] Given each distribution exhibits a Pareto tail, note that the top p-th percentile's share can be expressed as $(p/100)^{1-1/\xi}$ where ξ is the Pareto tail index. This expression shows when the Pareto distribution has a fatter

largest 1% firms ranked by employment, and 18.2% of all establishments are concentrated in the largest 1% of firms ranked by the number of establishments. All three distributions also exhibit substantial bunching at the bottom of the distribution at the discrete value of 1. In our dataset, 28% of firms have exactly one employee, 29% of firms have an average establishment size of one employee, and 96% of firms have exactly one establishment. Other discrete low values such as 2 and 3 exhibit bunching in the distribution, but less dramatically so.

Appendix E presents additional cross-sectional facts that we find in our dataset. There, we observe that the difference between small and large firms is due to both extensive and intensive margins, but among small firms, the difference is mostly explained by the intensive margin. The extensive margin exhibits a large contribution in the size differences and the growth of very large firms.

2.4 Time-series properties

Now we turn to the time-series changes in these distributions. We find notable changes in these distributions over our sample period. As depicted in Figure 2, average firm size increased from about 23 employees to over 25 employees since 1990. This fact is consistent with the rise in concentration in the U.S. economy documented by Autor et al. (2017).[16]

Figures 3 and 4 present the novel facts that we focus on in this paper. Figure 3 plots the average of the intensive margin and shows the intensive margin remained stationary (or somewhat declining) despite the increase in firm size over our sample period.[17] Therefore, our decomposition of firm size implies the average extensive margin must exhibit a strong upward trend, as is confirmed in Figure 4. The extensive margin grew from 1.2 in 1990 to over 1.5 in 2014. Accordingly, average firm size over 1990-2014 can be accounted for by the extensive-margin growth in the number of establishments.

This contrasting behavior between the intensive and extensive margins implies different forces are at work for these different components of firm growth.[18] To investigate what drives the increase in firm size, in particular along the extensive margin, we consider disaggregations by sector and size bins.

First, we examine the firm-size decomposition in the manufacturing, service, and agricultural sectors, and find that a significant increase in the number of service-sector establishments per

tail (lower ξ) firms at the top of a distribution employ a greater share of all resources relative to the rest of that distribution.

[16] Choi and Spletzer (2012) and Hathaway and Litan (2014) also document trends in firm size and establishment size, but do not analyze the intensive and extensive margins.

[17] Recent papers by Rinz (2018), Rossi-Hansberg et al. (2018), and Hershbein et al. (2019) document the diverging trends in national concentration and local concentration, which is analogous to the diverging trends in the average firm size and the intensive margin. Note, once again, that the average intensive margin is conceptually different from the average establishment size in the economy. One can compute the average establishment size from publicly-available data, but one needs to access the micro-level data to compute the average intensive margin.

[18] A recent paper by Argente et al. (2019) document that the sales of the individual products decline over time, and emphasize the introduction of new products as a source of the firm sales growth. The expansion of firms by adding establishments are also analyzed in the industrial organization literature, such as Holmes (2011), and international trade literature, such as Garetto et al. (2019).

Figure 2: Average firm size (number of workers)

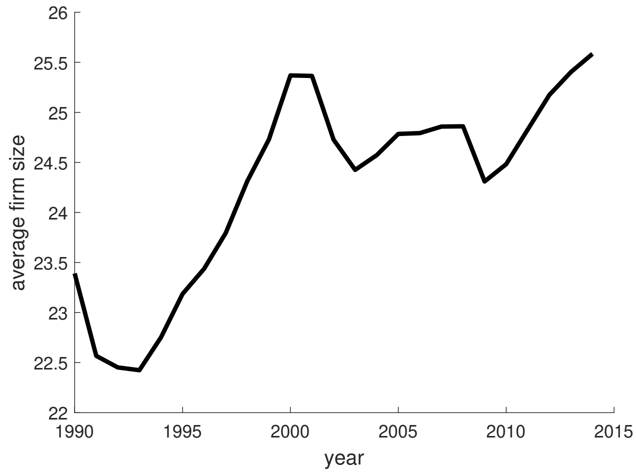


Figure 3: Average intensive margin

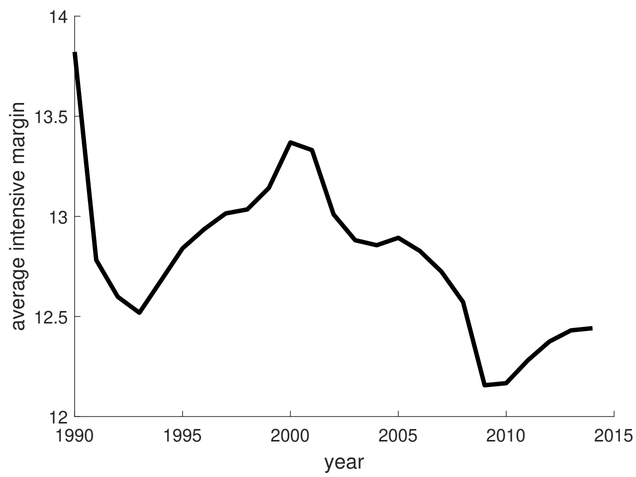
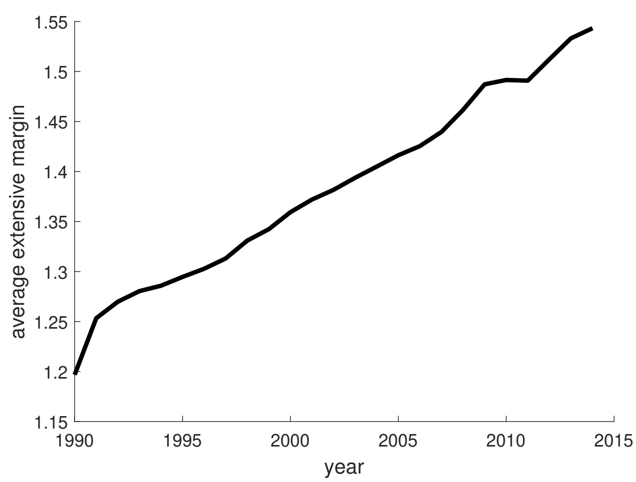


Figure 4: Average extensive margin



Source: Author's calculations of Quarterly Census of Employment and Wages microdata.

firm is the driving force for the economy-wide increase in average firm size. Figure 5(a) plots the evolution of average firm size in each sector, compared to the average firm size in 1990. We observe that all sectors excluding agriculture experienced an increase in average firm size over the sample period (1991-2014), but the service sector experienced the largest size increase. This observation is notable, because the service sector employs the majority of U.S. workers over this period. To account for sectoral firm size growth, we turn to the intensive and extensive margins plotted in Figures 5(c) and 5(e), respectively. Each sector's intensive margin exhibits a flat or slight downward trend similar to that in the overall economy. By strong contrast, the extensive margin for different sectors delivers the same message as the extensive margin in the overall economy, namely, that the growth in the number of establishments per firm accounts for the overall increase in average firm size across sectors and, by extension, the overall economy.

Next, we examine the firm-size decomposition conditional on firm size and find the establishment-driven growth in average firm size is concentrated in the economy's largest firms. Figure 5(b) calculates the average size within size bins. We see a pattern of spreading out: very small firms with one to four employees have tended to become smaller, whereas the average size of larger firms with 100 employees increased over time. If we examine the very right tail of firms with 5000 workers or more, firm size has been increasing over time since 1997, with a similar increase in firms that have 100 employees or more. The intensive margin does not exhibit an obvious relationship with firm size, as seen in Figure 5(d). None of the series have an increasing trend, and in fact, the overall time-series pattern looks similar between very small firms (1 to 4 employees) and very large firms (5,000 or more employees) except for a spike for very large firms in the early 2000s. By contrast, growth in the average number of establishments per firm exhibits very different trends between small and large firms, as shown in Figure 5(f). Very small firms are predominantly single-establishment firms over the entire sample period. Medium-size firms with 5 to 99 employees have had a modest increase in the number of establishments. Larger firms have had a startling increase in the number of establishments. On average, the firms with 5,000 or more employees had about four times more establishments in 2014 than in 1990. Thus, we conclude that a key mechanism that has generated the increase in firm size in recent years is expansion through the number of establishments in very large firms.

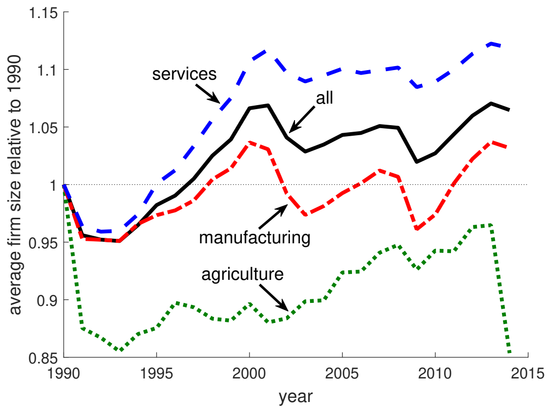
To see the behavior of the distribution at the right tail, we measure the slope of the upper percentiles of the firm-size distribution in Table 1.[19] Here, we include predicted values that are at or above the 95th percentile. The firm size distribution in 2014 has a slope that is close to (negative) 1, which indicates it has a very fat tail. Both the extensive margin and the intensive margin have steeper slopes than the employment distribution, which implies thinner tails. For the overall firm size, the right tail became thicker—looking at the 99 percentile and above, the tail index changed from -1.17 to -0.99 between 1995 and 2014. Table 1 indicates both the extensive and intensive margins contributed to this thickening of the slope over time.

Appendix E.3 describes additional time-series facts. There, we find the increase in the average

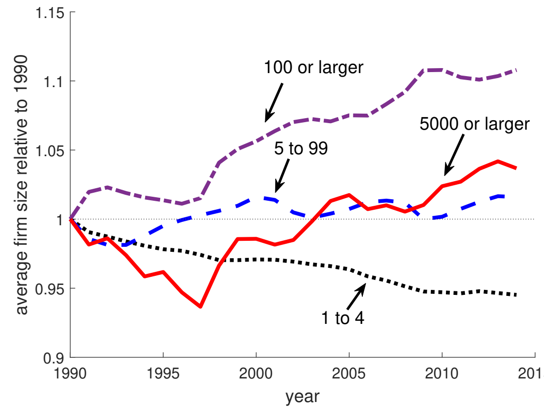
[19]For details on how we summarize our polynomial approximations with bivariate regressions, see Appendix D.2.4.

Figure 5: Average Firm Size, Intensive and Extensive Margins by Sector and Size Bins

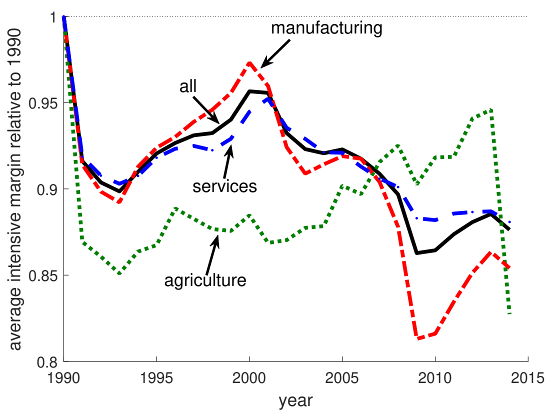
(a) Average firm size by sector (number of workers)



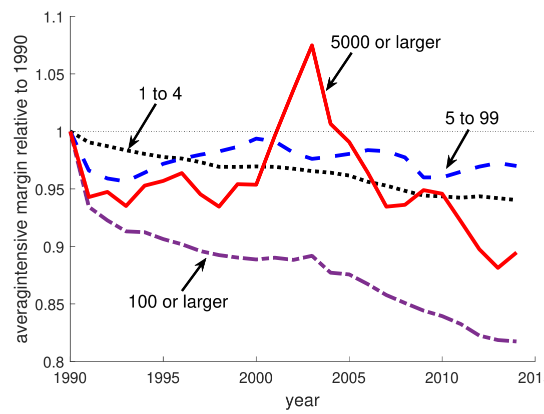
(b) Average firm size by size bin (number of workers)



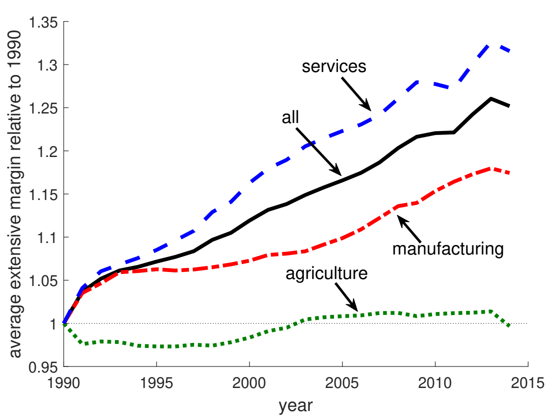
(c) Average intensive margin by sector



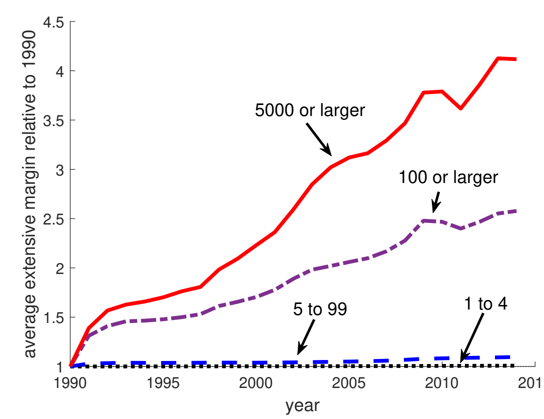
(d) Average intensive margin by size bin



(e) Average extensive margin by sector



(f) Average extensive margin by size bin



Source: Author's calculations of Quarterly Census of Employment and Wages microdata.

Table 1: Slope of the Size-Rank Relationship

	Firm size	Extensive	Intensive
	95th percentile and above		
1995	−1.10	−1.20	−1.35
2014	−0.99	−1.17	−1.32
	99th percentile and above		
1995	−1.17	−1.25	−1.39
2014	−0.99	−1.21	−1.24

Notes: Authors' calculations of of LEHD microdata. Linear regression of log outcome on log rank for fitted values at or above the 95th percentile from a polynomial approximation of microdata. Slopes correspond to March of the respective year.

extensive margin is more pronounced for older firms. This finding indicates the time trend is largely due to the change in the patterns of firm growth.

Overall, the preceding empirical documentation of firm growth shows the growth in average firm size between 1990 and 2014 is the result of high growth in creating new establishments, particularly by very large firms in the service sector. At the right tail, we observe increasing concentration in the firm-size distribution and the extensive-margin distribution between 1995 and 2014.[20]

3 Model

Motivated by the facts in the previous section, we construct a model of firm dynamics. Particularly, in Figure 1, the existence of Pareto tails in the intensive and extensive margins indicates Gibrat's Law holds for large firms. This finding imposes a theoretical structure on the driving forces behind the growth in both margins. We present a model in which intermediate-good firms make *endogenous* productivity improvements that lead to growth in both establishment size and the number of establishments. In Section 6, we use this model to quantitatively analyze the fundamental causes of the growth in firm size along the extensive margin and the increasing concentration of the firm-size distribution.

3.1 Model setting

Time is continuous. The representative consumer provides labor and consumes a final good. The final good is produced by combining differentiated intermediate goods.

[20]The comparison in Table 1, as well as the model estimation in Section 6, is between 1995 and 2014, instead of between 1990 and 2014, is due to the availability of the LEHD data.

3.2 Representative consumer

The consumer side is intentionally kept simple, because we focus mainly on firm growth. The utility function of the representative households is

$$U = \int_0^\infty e^{-\tilde{\rho}t} L(t) u(C(t)/L(t)) dt,$$

where $u(C(t)/L(t)) = (C(t)/L(t))^{1-\sigma}/(1-\sigma)$ for $\sigma > 0$ and $\sigma \neq 1$ or $u(C(t)/L(t)) = \log(C(t)/L(t))$, corresponding to $\sigma = 1$. The consumer consumes, owns firms, and supplies labor inelastically. The labor supply is given exogenously and grows at the rate $\gamma \geq 0$. Denoting the real interest rate as r, the Euler equation for the consumer is

$$\frac{\dot{C}(t)}{C(t)} = \frac{r - \rho}{\sigma}, \tag{1}$$

where $\rho \equiv \tilde{\rho} - \gamma$. Final-good output is used for consumption, firm investments in innovative activities, and firm fixed costs:

$$Y(t) = C(t) + R(t) + E(t),$$

where $Y(t)$ is the final goods output, $R(t)$ is total R&D of the incumbents, and $E(t)$ is the total entry cost.

3.3 Final-good producers

The final-good sector is perfectly competitive. The final good is produced from differentiated intermediate goods. Intermediate goods have different qualities, and a high-quality intermediate good contributes more to the final-good production. The production function for the final good is

$$Y(t) = \left(\int_{\mathcal{N}(t)} q_j(t)^\beta x_j(t)^{1-\beta} dj \right)^{\frac{1}{1-\beta}}, \tag{2}$$

where $x_j(t)$ is the quantity of intermediate good j, and $q_j(t)$ is its quality. $\mathcal{N}(t)$ is the set of actively-produced intermediate goods and $N(t) = |\mathcal{N}(t)|$ denotes the number of actively-produced intermediate goods. We assume $\beta \in (0, 1)$ so that the elasticity of substitution between differentiated goods $(1/\beta)$ is greater than one.

With the maximization problem

$$\max_{x_j(t)} \left(\int_{\mathcal{N}(t)} q_j(t)^\beta x_j(t)^{1-\beta} dj \right)^{\frac{1}{1-\beta}} - \int_{\mathcal{N}(t)} p_j(t) x_j(t) dj,$$

the inverse demand function for the intermediate good j is

$$p_j(t) = Y(t)^{\beta} \left(\frac{q_j(t)}{x_j(t)} \right)^{\beta}. \tag{3}$$

3.4 Intermediate-good producers

Production: The intermediate-good sector is monopolistically competitive. Each intermediate good is produced by one firm. A firm can potentially produce many intermediate goods. A firm can add a new intermediate good to its portfolio by investing in R&D that generates an *external innovation*. It can also increase the quality of the intermediate goods that it already produces by investing in R&D that generates an *internal innovation*. A new firm can enter the market by innovating its first product. In our calibration, we map one intermediate good in the model to one establishment in the data.

We assume the intermediate goods are produced only by labor. This production process is the only place in the entire economy that uses labor as an input, thus allowing us to map the employment dynamics of the intermediate-good sector to our data analysis in Section 2.[21] The production function for intermediate good j is

$$x_j(t) = A(t)\ell_j(t), \tag{4}$$

where $A(t)$ is exogenous labor productivity that grows at rate θ.

Given the final-good producer's demand for its output (3), the intermediate goods producer's profit maximization results in standard optimal pricing with a markup over marginal cost:

$$p_j(t) = \frac{1}{1-\beta}\frac{w(t)}{A(t)}.$$

The optimal price, together with (3) and (4), implies the labor demand is proportional to $q_j(t)$ given the aggregate variables

$$\ell_j(t) = (1-\beta)^{\frac{1}{\beta}} w(t)^{-\frac{1}{\beta}} A(t)^{\frac{1-\beta}{\beta}} Y(t) q_j(t),$$

and is increasing in $A(t)$ and $Y(t)$ and is decreasing in $w(t)$. Using the normalized wage defined as $\bar{w}(t) \equiv w(t)/(A(t)Y(t)^{\frac{\beta}{1-\beta}})$, we rewrite the labor demand as

$$\ell_j(t) = \frac{(1-\beta)^{\frac{1}{\beta}} \bar{w}(t)^{-\frac{1}{\beta}}}{A(t)Y(t)^{\frac{\beta}{1-\beta}}} q_j(t). \tag{5}$$

Because firms (and establishments) differ only in their level of quality $q_j(t)$, employment per establishment varies proportionally to $q_j(t)$ in the cross section. Similarly, profit is also proportional

[21] A similar idea of mapping the employment process to the productivity process is employed by Hopenhayn and Rogerson (1993), Garca-Macia et al. (2019), and Mukoyama and Osotimehin (2019).

to $q_j(t)$:

$$\pi_j(t) = \bar{\pi}(t) q_j(t),$$

where

$$\bar{\pi}(t) \equiv \beta(1-\beta)^{\frac{1-\beta}{\beta}} \bar{w}(t)^{\frac{\beta-1}{\beta}}. \tag{6}$$

Innovation: Innovations are carried out through R&D activity. The input for R&D is in final goods. For an existing intermediate-good firm, two kinds of innovations are possible: internal innovation and external innovation.

Internal innovation raises the quality of the goods that a firm already produces. The firm-level total intensity of internal innovation is denoted by $Z_{I,j}(t)$. The innovation intensity per good is $z_{I,j}(t) \equiv Z_{I,j}(t)/n_j(t)$, where $n_j(t)$ is the (discrete) number of goods firm j produces. Then, the quality improves according to the law of motion:

$$\frac{dq_j(t)}{dt} = z_{I,j}(t) q_j(t).$$

Here we index q only by j because, as we explain momentarily, the quality of all goods produced by firm j is always the same within the firm.

We assume different firms can have different costs for innovation. In particular, we partition firms into different (finite) types, and assume different types have different costs for innovation. We will detail later how types evolve over time. We denote the number of types by T and index the types by τ. The R&D cost for internal innovation is assumed to be $\mathbf{R}_I^\tau(Z_{I,j}(t), n_j(t), q_j(t))$. As in Klette and Kortum (2004), we assume the R&D cost function $\mathbf{R}_I^\tau(Z_{I,j}(t), n_j(t), q_j(t))$ exhibits constant returns to scale with respect to $Z_{I,j}(t)$ and $n_j(t)$. Then, the R&D cost per good can be denoted as

$$R_I^\tau(z_{I,j}(t), q_j(t)) \equiv \mathbf{R}_I^\tau(z_{I,j}(t), 1, q_j(t)) = \frac{\mathbf{R}_I^\tau(Z_{I,j}(t), n_j(t), q_j(t))}{n_j(t)}.$$

We further assume

$$R_I^\tau(z_{I,j}(t), q_j(t)) = h_I(z_{I,j}(t)) q_j(t)$$

for a strictly convex function $h_I^\tau(\cdot)$.

External innovation adds brand-new intermediate goods to the production portfolio of the firm. We assume the new good has the same quality as the average quality of the goods produced by that firm. Thus, all products that firm j produces always have the same quality. The total intensity of external innovation is denoted by $Z_{X,j}(t)$. The innovation intensity per good is $z_{X,j}(t) \equiv Z_{X,j}(t)/n_j(t)$. The R&D cost for external innovation is assumed to be $\mathbf{R}_X^\tau(Z_{X,j}(t), n_j(t), q_j(t))$, which is assumed to be constant returns to scale with respect to $Z_{X,j}(t)$ and $n_j(t)$. Once again, we can denote the cost per good as

$$R_X^\tau(z_{X,j}(t), q_j(t)) \equiv \mathbf{R}_X^\tau(z_{X,j}(t), 1, q_j(t)) = \frac{\mathbf{R}_X^\tau(Z_{X,j}(t), n_j(t), q_j(t))}{n_j(t)},$$

and we assume

$$R_X^\tau(z_{X,j}(t), q_j(t)) = h_X^\tau(z_{X,j}(t))q_j(t)$$

for a strictly convex function $h_X^\tau(\cdot)$.[22]

Dynamic programming problem: We assume firms transition between different types from τ to τ' with Poisson transition rates $\lambda_{\tau\tau'}$. Each establishment depreciates (is forced to exit) with the Poisson rate δ_τ. We also impose an exogenous exit shock at the firm level. Let d_τ be the Poisson rate of the firm exit shock for a type-τ firm. We omit time notation here, because all variables and functions are constant over time along the balanced-growth path (BGP) that we construct.

Each firm is a collection of n establishments that are each characterized by a quality level, $\{q_j\}_{j=1}^n$. Let $\mathbf{V}_\tau$ denote the value function of the firm, such that the Hamilton-Jacobi-Bellman (HJB) equation for the firm is

$$r\mathbf{V}_\tau(\{q_j\}) = \max_{z_{I,j}, z_{X,j}} \sum_{j=1}^n \Pi(\{q_j\}, z_{I,j}, z_{X,j}) + \sum_{\tau'} \lambda_{\tau\tau'}(\mathbf{V}_{\tau'}(\{q_j\}) - \mathbf{V}_\tau(\{q_j\})) + \dot{\mathbf{V}}_\tau(\{q_j\}),$$

where the return function is[23]

$$\begin{aligned}\Pi(\{q_j\}, z_{I,j}, z_{X,j}) \equiv{} & \bar{\pi}q_j - R_I^\tau(z_{I,j}, q_j) - R_X^\tau(z_{X,j}, q_j) \\ & + z_{I,j}\frac{\partial \mathbf{V}_\tau(\{q_j\})}{\partial q_j}q_j + z_{X,j}(\mathbf{V}_\tau(\{q_{-j}\} \cup \{q_j, q_j\}) - \mathbf{V}_\tau(\{q_j\})) \\ & - \delta_\tau \mathbf{V}_\tau(\{q_{-j}\}) - d_\tau \mathbf{V}_\tau(\{q_j\}).\end{aligned}$$

In this expression, $\sum_{j=1}^n \bar{\pi}q_j$ is the total profit obtained from the establishments, and $R_I^\tau(\cdot), R_X^\tau(\cdot)$ are the previously discussed functions governing investment in internal and external innovations, that yield the intensive and extensive expansion rates, (z_I, z_X), respectively.[24]

Because of separability, the value function for the firm is the sum of the value functions across establishments,

$$\mathbf{V}_\tau(\{q_j\}) = \sum_{j=1}^n V_\tau(q_j),$$

[22] A major departure from Luttmer (2011) is that we allow for internal innovation. Internal innovation allows us to capture the characteristics of intensive-margin growth; we have shown in Section 2 that establishment size grows over time as firms age.

[23] $\{q_{-j}\}$ denotes the initial set of qualities excluding q_j.

[24] From (3) and (4), the firm-level revenue can be written as $\sum_{j=1}^n Y(t)^\beta (q_j(t))^\beta (A(t)\ell_j(t))^{1-\beta}$. When a firm has n identical establishments with identical product quality $q(t)$ (this is the case in equilibrium) and total labor $\ell(t)$, the total revenue is $Y(t)^\beta (q(t))^\beta (A(t)\ell(t))^{1-\beta} n^\beta$. For a given n, the revenue function exhibits decreasing returns to ℓ. For firms with small values of n, such as $n = 1$ or $n = 2$, we expect the firm to behave as a decreasing-returns producer, given that n tend to remain fixed. When n is large, however, a large firm has both large ℓ and n, because q and n tend to grow together. For these firms, with ℓ and n both interpreted as production factors, the revenue function can be viewed to have higher returns to scale. Consistent with this intuition, Dinlersoz et al. (2018) find privately-owned firms exhibit financing behaviors associated with decreasing-returns-to-scale production functions, whereas publicly-listed firms do not. The latter typically have more establishments.

and the establishment-level HJB equation of a type-τ establishment is

$$rV_\tau(q) = \max_{z_I, z_X} \left[\begin{array}{c} \bar{\pi}q - R_I^\tau(z_I, q) - R_X^\tau(z_X, q) + z_I \frac{\partial V_\tau(q)}{\partial q} q \\ + z_X V_\tau(q) - (\delta_\tau + d_\tau) V_\tau(q) + \sum_{\tau'} \lambda_{\tau\tau'} (V_{\tau'}(q) - V_\tau(q)) \end{array} \right],$$

where $V_\tau(q)$ is the value of type-τ establishment with quality q and $\dot{V}_\tau(q)$ is the time derivative of $V_\tau(q)$ function.

As in Mukoyama and Osotimehin (2019), $V_\tau(q)$ can be shown to be linearly homogeneous in q along the BGP. That is, $V_\tau(q) = v_\tau q$ for a constant v_τ. The HJB equation above can be normalized to

$$rv_\tau = \max_{z_I, z_X} \left[\bar{\pi} - h_I^\tau(z_I) - h_X^\tau(z_X) + (z_I + z_X - \delta_\tau - d_\tau) v_\tau + \sum_{\tau'} \lambda_{\tau\tau'} (v_{\tau'} - v_\tau) \right], \tag{7}$$

where $\bar{\pi}$ is given by (6). Accordingly, the HJB equation (7) implies that the choice of innovation intensities (z_I, z_X) is a function of the firm type only. We denote the decision rules as (z_I^τ, z_X^τ).[25]

Entry: An intermediate firm can enter the market by creating a new product. A new firm draws its type from an exogenous distribution, where m_τ is the probability that an entrant draws the type τ. Given a type τ, the entrant draws its initial *relative quality* $\hat{q}$ from a distribution $\Phi_\tau(\hat{q})$. We assume this relative quality $\hat{q}$ is equal to $q(t)/Q(t)$, where

$$Q(t) \equiv \frac{1}{N(t)} \int_{\mathcal{N}(t)} q_j(t) dj$$

is the average quality of intermediate goods. The firm's value of entry $V^e(t)$ is thus

$$V^e(t) = \sum_\tau m_\tau \int V_\tau(\hat{q}Q(t)) d\Phi_\tau(\hat{q}).$$

We assume that any potential entrant can pay a cost $\phi Q(t)$, denominated in final goods, to begin production. Therefore, the free-entry condition is: $V^e(t) = \phi Q(t)$. By defining the value of entry relative to average product quality, $v^e \equiv V^e(t)/Q(t)$, we can rewrite the value of entry as

$$v^e = \sum_\tau m_\tau v_\tau \int \hat{q} d\Phi_\tau(\hat{q}) \tag{8}$$

and the free-entry condition as

$$v^e = \phi. \tag{9}$$

Note that once r is given, we can find a value of $\bar{w}$ that satisfies the free entry condition (9). Let the number of entrants at time t be $\mu_e N(t)$, where μ_e is a constant along the BGP.

[25] Recall that, from (6), $\bar{\pi}$ is a function of $\bar{w}$ only. Thus, given r and $\bar{w}$, the equation (7) and the first-order conditions can solve for v_τ, z_I^τ, and z_X^τ.

3.5 Balanced-growth equilibrium

A *competitive equilibrium* of this economy is a wage $w(t)$, a consumer allocation $(C(t), R(t), E(t))$, a final-good-producer allocation $(Y(t), \{x_j(t)\}_{j\in\mathcal{N}(t)})$, an allocation for intermediate-good producers $(\{\ell_j(t), q_j(t), p_j(t), z_{I,j}(t), z_{X,j}(t), n_j(t)\}_{j\in\mathcal{N}(t)})$, and a value of entry $V_e(t)$ such that at each instant, (i) consumers optimize, (ii) the final-good producers' allocation solves its profit maximization problem, (iii) the intermediate-good producers' allocations solve their profit-maximization problem, (iv) the free-entry condition holds, (v) the final-good market clears: $Y(t) = C(t) + R(t) + E(t)$, and (vi) the labor market clears: $L(t) = \int_{\mathcal{N}_t} \ell_j(t)dj$.

We now construct a *balanced-growth equilibrium* of this economy. Assume the population $L(t)$ grows at an exogenous rate γ. Furthermore, let aggregate quality $Q(t)$ grow at a constant rate ζ and the number of establishments $N(t)$ (and the number of firms) grow at a constant rate η. Denote the growth rate of final output $Y(t)$ by g. Along a BGP, the growth rates of $Y(t)$, $C(t)$, $R(t)$, and $E(t)$ must all be equal. Thus, the Euler equation (1) requires $\dot{C}(t)/C(t) = g$ because $C(t)$ grows at the same rate as $Y(t)$. This fact implies $r = \rho - \sigma g$ along the BGP.

The quality-invariant component of profit $\bar{\pi}(t)$ in (6) is constant along the BGP. Therefore, $w(t)$ must grow at the same rate as $A(t)Y(t)^{\frac{\beta}{1-\beta}}$. Given that $Y(t)$ grows at rate g and $A(t)$ grows at rate θ, $w(t)$ must grow at the rate $\beta g/(1-\beta) + \theta$. Because labor income of the representative consumer must grow at the same rate as consumption, and because population growth is γ, the following relationship must hold:

$$g = \gamma + \theta + \frac{\beta}{1-\beta} g, \tag{10}$$

which implies output growth of $g = (1-\beta/(1-\beta))^{-1}(\gamma+\theta)$. Similarly to Luttmer (2011), the output growth rate is dictated by the population growth rate γ and the intermediate-good productivity growth rate θ.[26]

4 Characterization of the model

In this section, we analytically characterize aggregate growth and firm-level decisions in the model. These characterizations will be useful for understanding how the model draws identification from the data, as well as anchoring the interpretation of quantitative experiments in Section 6. We detail the characterization of distributions in Section 5.

4.1 Output growth

Output growth derives from firm-level investments in internal and external innovations. First, we examine a decomposition of output growth into the extensive- and intensive-margin growth rates, η and ζ, at an aggregate level. Using the labor-market-clearing condition $L(t) = \int_{\mathcal{N}_t} \ell_j(t)dj$ along

[26] In the terminology of Jones (1995), our model exhibits a "semi-endogenous" growth. In Appendix C.2, we explore the implication of an alternative model that exhibits fully endogenous growth.

with firm labor demand in equation (5) yields the following expression:

$$\bar{w}(t)L(t) = (1-\beta)\left(\frac{N(t)Q(t)}{A(t)Y(t)^{\frac{\beta}{1-\beta}}}\right)^{\beta} L(t)^{1-\beta}. \tag{11}$$

Because the normalized wage $\bar{w}(t)$ does not grow along the BGP, equation (11) implies a decomposition of output growth into the growth of the intensive and extensive margins,[27]

$$g = \eta + \zeta. \tag{12}$$

Because final output growth can be decomposed into the extensive and intensive margins of firm growth, it has a natural interpretation as the aggregate outcome of disaggregated firm behavior. First, consider the extensive margin. The total number of establishments of type τ is $N_\tau(t) \equiv M_\tau N(t)$, where $M_\tau \in [0,1]$ is the share of type τ establishments, and satisfies $\sum_\tau M_\tau = 1$. The law of motion for $N_\tau(t)$ is

$$\dot{N}_\tau(t) = z_X^\tau N_\tau(t) - (\delta_\tau + d_\tau)N_\tau(t) + \mu_e m_\tau N(t) - \sum_{\tau'\neq\tau} \lambda_{\tau\tau'} N_\tau(t) + \sum_{\tau'\neq\tau} \lambda_{\tau'\tau} N_{\tau'}(t).$$

The first term is the increase in the number of establishments due to external innovation, the second term is the loss of establishments due to exit, the third term accounts for product entry, and the fourth and fifth terms capture the product-number evolution due to changes in firm types. On the BGP, $N_\tau(t)$ grows at rate η for all τ, and thus this equation can be rewritten as:[28]

$$\eta = z_X^\tau - (\delta_\tau + d_\tau) + \mu_e \frac{m_\tau}{M_\tau} - \sum_{\tau'\neq\tau} \lambda_{\tau\tau'} + \sum_{\tau'\neq\tau} \lambda_{\tau'\tau} \frac{M_{\tau'}}{M_\tau}, \tag{13}$$

Next, consider the intensive margin. Define the average quality of type τ firms as

$$Q_\tau(t) \equiv \frac{1}{N_\tau(t)} \int_{\mathcal{N}_\tau(t)} q_j(t)dj,$$

where $\mathcal{N}_\tau(t)$ is the set of actively produced goods by type-τ firms. Further define s_τ as the quality share of type τ firms by

$$s_\tau \equiv M_\tau Q_\tau(t)/Q(t), \tag{14}$$

[27] Equation (11) implies the growth relationship $\gamma = \eta + \zeta - \theta - (\beta/(1-\beta))g$, which is combined with equation (10) to yield equation (12).

[28] The growth rate of the total number of establishments can also be written as the weighted sum of the growth rates of the number of type-τ establishments and the entry rate,

$$\eta = \sum_\tau M_\tau [z_X^\tau - (\delta_\tau + d_\tau)] + \mu_e,$$

which is found by multiplying M_τ to both sides of (13) and summing over τ.

which satisfies

$$\sum_{\tau} s_\tau = 1. \tag{15}$$

On the BGP, s_τ is constant, which implies $Q_\tau(t)$ has to grow at the same rate as $Q(t)$ for all types τ.

Finally, using $g = \eta + \zeta$ defines aggregate output growth g as a function of firm-level innovations and shocks,

$$g = \left[z_I^\tau + z_X^\tau - (\delta_\tau + d_\tau)\right] + \mu_e \frac{m_\tau}{s_\tau} \int \hat{q} d\Phi(\hat{q}) - \sum_{\tau' \neq \tau} \lambda_{\tau\tau'} + \sum_{\tau' \neq \tau} \lambda_{\tau'\tau} \frac{s_{\tau'}}{s_\tau}. \tag{16}$$

The first term is the incumbent firms' contribution to g, the second term characterizes entrants' contribution to output growth, and the final terms capture the impact of changing firm types.[29]

4.2 Properties of the model with one type

To highlight the mechanisms underlying the model, here we consider the model with one type (omitting the τ subscript) and provide a sharp characterization of firm-level decisions and general equilibrium outcomes. Although our quantitative experiments in Section 6 utilize a specification with two types, the results with one type deliver essentially the same qualitative relationships.

First, the one-type growth rates for the intensive margin, extensive margin, and aggregate output are easily obtained from the previous section,

$$\zeta = z_I - \mu_e \left(1 - \int \hat{q} d\Phi(\hat{q})\right) \tag{17}$$

$$\eta = z_X - \delta - d + \mu_e \tag{18}$$

$$g = z_I + z_X - \delta - d + \mu_e \int \hat{q} d\Phi(\hat{q}). \tag{19}$$

To determine these growth rates, we must characterize firm-level decisions. To do so, consider the HJB equation of the firm,

$$(\rho + \sigma g)v = \bar{\pi} - h_I(z_I) - h_X(z_X) + (z_I + z_X - \delta - d)v,$$

whose first-order conditions imply innovation intensities z_I and z_X are solely determined by the firm's value v and the innovation-cost functions $h_I(\cdot)$ and $h_X(\cdot)$. We can use the free entry conditions (8) and (9) to obtain the firm's value from the entry cost, $v = \phi / \int \hat{q} d\Phi(\hat{q})$. Thus, the only remaining

[29] A simpler expression for g can be found by multiplying s_τ on both sides of (16) and summing across τ,

$$g = \sum_{\tau} s_\tau [z_I^\tau + z_X^\tau - (\delta_\tau + d_\tau)] + \mu_e \int \hat{q} d\Phi(\hat{q}).$$

Appendix A contains a derivation for equation (16).

equilibrium object we need for determining growth rates (ζ, η) is the entry rate, μ_e. The entry rate can be obtained as the solution to equations (10) and (19),

$$\mu_e = \frac{1}{\int \hat{q} d\Phi(\hat{q})} \left(\frac{\gamma + \theta}{1 - (\beta/(1-\beta))} - \Big(z_I + z_X - \delta - d \Big) \right), \tag{20}$$

where we assume $\beta < 1/2$ to ensure positive entry in equilibrium $(\mu_e > 0)$. Finally, the value of $\bar{\pi}$ is pinned down by the HJB after obtaining values for (v, z_I, z_X).

Given this characterization, the following comparative static exercises are straightforward and are therefore presented without proof.

Proposition 1 *Consider a BGP of the one-type economy. The following comparative statics hold:*

(i) ***Entry Costs:*** *An increase in the entry cost, ϕ, generates an increase in innovation intensities (z_I, z_X) and a decrease in the entry rate (μ_e). These effects exactly offset and g remains constant.*

(ii) ***Innovation Costs:*** *Suppose innovation-cost functions take the form $h_i(z_i) = \chi_i z_i^{\psi}$, where parameters satisfy $\chi_i > 0$ and $\psi > 1$ for $i \in \{I, X\}$. Then a decrease in χ_I increases z_I but does not affect z_X. A decrease in χ_X increases z_X but does not affect z_I. The entry rate (μ_e) decreases when either cost χ_I or χ_X decreases.*

(iii) ***Technology:*** *An increase in population growth γ, productivity growth θ, or exit rates (δ, d) does not affect innovation intensities (z_I, z_X) but does generate higher output growth g and firm entry μ_e. Changes in the parameter governing the elasticity of substitution, β, do not affect z_I or z_X, although they influence g and μ_e.*

(iv) ***Preferences:*** *Changes in the preference parameters (σ, ρ) do not affect z_I, z_X or g.*

From Proposition 1, one key prediction of the model is a trade-off between incumbent investment in innovative activity and firm entry. If the entry cost ϕ increases, the free-entry condition implies a firm needs to receive higher lifetime compensation for incurring the increased initial cost of creating a new product for the market. Therefore, conditional on entering, the firm will invest in innovative activity at a higher rate, offsetting the high entry cost, and subsequently grow to a larger size along both the intensive and extensive margins.[30]

However, the balanced-growth condition in equation (10) requires the wage not to grow too fast relative to productivity and final output. This condition makes it clear that there are limits to growth due to population and technology dynamics. In particular, incumbent firms' increased innovative activity diminishes the incentive of new firms to enter the market due to labor scarcity. Similarly, exit encourages firm entry because exit frees up resources for production and entrants have the incentive to engage in product innovation if there are resources with which they can grow

[30]In the two-type model of Section 6, because χ_I and χ_X may vary by firm type, innovative activity can have more heterogeneity. To match the empirical observations in Section 2.1, we indeed find innovation-cost parameters vary by firm type.

and recover their initial investment. In this sense, the economy exhibits a crowding-out of entrants by innovative incumbents.

Finally, note that if aggregate output growth were (partially) endogenized, for example, through an externality that leads innovative activity to increase productivity growth, the trade-off between incumbent innovative activity and firm entry would be relaxed. As a result, the model can infer the extent to which aggregate output growth is endogenous by using data on entry over time, given a set of identified innovation costs. We pursue this extension of the model as a robustness exercise in Appendix C.2.

5 Distributions of firm sizes and establishment sizes

The properties of our model allow us to analyze the firm-size distribution from two different margins. First, the number of establishments per firm evolves through external innovation and exit shocks. Second, the size of each establishment evolves through internal innovation.[31] The first margin exactly corresponds to the extensive margin in Section 2. The second margin does not directly correspond to the intensive margin in Section 2 (because economy-wide establishment size distribution is not the same as economy-wide intensive margin distribution), but is closely related.

Before analyzing the details, we first note one general property of the model. The model assumptions imply establishments are homogeneous within a firm. A firm starts with one establishment, and whenever it expands the number of establishments, a new establishment inherits the same quality as the existing establishments. The intensity of internal innovation, z_I^τ, is common across establishments within a firm, although it may change over time. These two facts mean that establishments share common quality within a firm at any point in time with the type switching. This property also implies that establishment sizes are also common within a firm. Although in reality establishment sizes are not the same within a firm, we view this assumption as a useful simplification that affords us sharp analytical characterizations. In a balanced-growth equilibrium, the number of firms grows at the same rate as the number of establishments $N(t)$. The distribution of the number of establishments per firm, $n(t)$, is stationary (in shares), and the distribution of establishment's relative quality, $q(t)/Q(t)$, or size, is also stationary, despite the fact that $Q(t)$ grows exponentially over time. In the following, we provide further characterizations of the size distributions of establishments and firms. We assume the economy is on a stationary BGP.

For the characterization of the distributions, we use the following mathematical notations: for two strictly positive functions f, g defined over $(0, x^*)$, where $x^* \in \mathbb{R}^* \cup \{+\infty\}$,

$$f(x) \sim_{x \to x^*} g(x)$$

[31] Note the existence of these two margins is the major departure from Klette and Kortum (2004) and Luttmer (2011). In these papers, establishments are homogeneous and each establishment does not grow, so the only relevant innovation is external innovation.

if $\lim_{x\to x^*}\frac{f(x)}{g(x)}=1$;[32] and

$$f(x)\propto_{x\to x^*} g(x)$$

if $f(x)\sim_{x\to x^*} ag(x)$ for some $a>0$. A random variable $\mathbf{X}$ defined over $\mathbb{R}^*$ has a Pareto tail with tail index $\xi>0$ if $\Pr(\mathbf{X}>x)\propto_{x\to\infty} x^{-\xi}$, or equivalently

$$\lim_{x\to\infty} x^{\xi}\Pr(\mathbf{X}>x)=a$$

for some $a>0$. The distribution has a thin tail if

$$\lim_{x\to\infty} x^{\xi}\Pr(\mathbf{X}>x)=0$$

for any $\xi>0$. For notational convenience, we assign tail index ∞ to distributions with a thin tail.

5.1 General characterizations

When multiple types of firms exist, characterizing the joint distribution of $(n,\hat{q})$ is substantially more complex than the one-type case. Our approach here is to look at two separate margins: the distribution of the number of establishments per firm, summarized by $\bar{\mathcal{M}}_\tau(n)$, which is the measure of type-τ firms with n establishments divided by $N(t)$ (i.e., $N(t)$-normalized measure), and the distribution of establishment quality relative to average quality, $q(t)/Q(t)$, a measure of establishment size.[33] We denote the fraction of type-τ establishments with $q(t)/Q(t)\geq\hat{q}$ as $\bar{\mathcal{H}}_\tau(\hat{q})$.

The distribution of establishment numbers, $\{\bar{\mathcal{M}}_\tau(n)\}_{n=1}^{\infty}$, can be characterized by the following difference equations:

$$\begin{aligned} 0 = \;& -(z_X^\tau+\delta_\tau+d_\tau+\eta)\bar{\mathcal{M}}_\tau(1)+2\delta_\tau\bar{\mathcal{M}}_\tau(2)+\mu_e m_\tau \\ & -\textstyle\sum_{\tau'\neq\tau}\lambda_{\tau\tau'}\bar{\mathcal{M}}_\tau(1)+\sum_{\tau'\neq\tau}\lambda_{\tau'\tau}\bar{\mathcal{M}}_{\tau'}(1) \end{aligned} \tag{21a}$$

for each τ, and

$$\begin{aligned} 0 = \;& -(n(z_X^\tau+\delta_\tau)+d_\tau+\eta)\bar{\mathcal{M}}_\tau(n)+(n+1)\delta_\tau\bar{\mathcal{M}}_\tau(n+1)+(n-1)z_X^\tau\bar{\mathcal{M}}_\tau(n-1) \\ & -\sum_{\tau'\neq\tau}\lambda_{\tau\tau'}\bar{\mathcal{M}}_\tau(n)+\sum_{\tau'\neq\tau}\lambda_{\tau'\tau}\bar{\mathcal{M}}_{\tau'}(n) \end{aligned} \tag{21b}$$

for $n>1$ and each τ.

The distribution of establishment-level relative quality, which is proportional to establishment size, $\bar{\mathcal{H}}_\tau(\hat{q})$, is governed by the following Kolmogorov forward equation:

$$\begin{aligned} (z_I^\tau-\zeta)\hat{q}\frac{d\bar{\mathcal{H}}_\tau(\hat{q})}{d\hat{q}} = \;& -(\delta_\tau+d_\tau+\eta-z_X^\tau)\bar{\mathcal{H}}_\tau(\hat{q})+\mu_e\frac{m_\tau}{M_\tau}(1-\Phi(\hat{q})) \\ & -\sum_{\tau'\neq\tau}\lambda_{\tau\tau'}\bar{\mathcal{H}}_\tau(\hat{q})+\sum_{\tau'\neq\tau}\lambda_{\tau'\tau}\frac{M_{\tau'}}{M_\tau}\bar{\mathcal{H}}_{\tau'}(\hat{q}). \end{aligned} \tag{22}$$

[32] This $\sim$ notation follows the regular variation literature (Bingham et al. (1987)).
[33] On a BGP, these normalized measures are constant and $\sum_{\tau,n} n\bar{\mathcal{M}}_\tau(n)=1$.

Lastly, for the distribution over firm size, recall all establishments in a firm grow at the same rate and hence have the same size. Therefore, we just need to keep track of the joint distribution of establishment number and establishment size in order to study the firm-size distribution. That is, the distribution of firm size is some "convolution" of the distribution of the number of establishments per firm and the distribution of establishment size and can be derived as follows. Let $\mathcal{M}_\tau(n, \hat{q})$ be the normalized measure of type-τ firms with n establishments and $q(t) \geq \hat{q}Q(t)$. Then

$$\begin{aligned}(z_I^\tau - \zeta)\hat{q}\frac{d\mathcal{M}_\tau(1,\hat{q})}{d\hat{q}} =& -(z_X^\tau + \delta_\tau + d_\tau + \eta)\mathcal{M}_\tau(1,\hat{q}) \\ &+ 2\delta_\tau \mathcal{M}_\tau(2,\hat{q}) + \mu_e m_\tau (1 - \Phi(\hat{q})) \\ &+ \sum_{\tau' \neq \tau} \lambda_{\tau'\tau}\mathcal{M}_{\tau'}(1,\hat{q}) - \sum_{\tau' \neq \tau} \lambda_{\tau\tau'}\mathcal{M}_\tau(1,\hat{q}) \end{aligned} \tag{23a}$$

for each τ, and

$$\begin{aligned}(z_I^\tau - \zeta)\hat{q}\frac{d\mathcal{M}_\tau(n,\hat{q})}{d\hat{q}} =& -(n(z_X^\tau + \delta_\tau) + d_\tau + \eta)\mathcal{M}_\tau(n,\hat{q}) \\ &+ (n+1)\delta_\tau \mathcal{M}_\tau(n+1,\hat{q};t) + (n-1)z_X^\tau \mathcal{M}_\tau(n-1,\hat{q}) \\ &+ \sum_{\tau' \neq \tau} \lambda_{\tau'\tau}\mathcal{M}_{\tau'}(n,\hat{q}) - \sum_{\tau' \neq \tau} \lambda_{\tau\tau'}\mathcal{M}_\tau(n,\hat{q}) \end{aligned} \tag{23b}$$

for $n > 1$ and each τ. The detailed derivations of these equations are presented in Appendix B.1.

5.2 Distributions for one-type economy

When only one firm type exists, firm growth is governed by three endogenous variables: z_I, z_X, and μ_e. Note that, in this case, our model assumptions imply that for a given firm, the quality (and therefore the size) of each establishment grows at a deterministic rate z_I that is common across all firms. The average quality $Q(t)$ grows at the rate ζ given in (17). Thus, the quality of the establishments in a firm that entered at time t_0 and whose initial draw of the normalized quality is $\hat{q}Q(t_0)$ can be represented as (denoting it by $q_{t_0}(t)$)

$$q_{t_0}(t) = \hat{q}Q(t_0)e^{z_I(t-t_0)} = \hat{q}Q(t)e^{(z_I-\zeta)(t-t_0)}.$$

From the labor demand (5) and the labor market equilibrium condition, it is straightforward to show that along the balanced-growth equilibrium, the relative labor demand $\ell(t)/L(t)$ of a particular establishment with quality $q_{t_0}(t)$ is equal to $q_{t_0}(t)/(N(t)Q(t))$. Therefore, the cross-sectional distribution of establishment size at a given time t is the same as the distribution of $\hat{q}e^{(z_I-\zeta)(t-t_0)}$. Denoting the time-$t$ number of establishments for a firm that starts at time t_0 as $n_{t_0}(t)$ (note that $n_{t_0}(t)$ is stochastic as the external innovation is random), the (relative) firm-size distribution follows the distribution of $n_{t_0}(t)\hat{q}e^{(z_I-\zeta)(t-t_0)}/N(t)$.

Equation (22) becomes

$$(z_I - \zeta)\hat{q}\frac{d\bar{\mathcal{H}}(\hat{q})}{d\hat{q}} = -(\delta + d + \eta - z_X)\bar{\mathcal{H}}(\hat{q}) + \mu_e(1 - \Phi(\hat{q})).$$

Let us use the change of variables $p \equiv \log(\hat{q})$ and $\tilde{\mathcal{H}}(p) \equiv \bar{\mathcal{H}}(\exp(p))$ to rewrite this equation as

$$(z_I - \zeta)\frac{d\tilde{\mathcal{H}}(p)}{dp} = -(\delta + d + \eta - z_X)\tilde{\mathcal{H}}(p) + \mu_e(1 - \Phi(\exp(p))).$$

This equation is a first-order ODE that has a general solution:

$$\tilde{\mathcal{H}}(p) = e^{\frac{\delta+d+\eta-z_X}{z_I-\zeta}(\underline{p}-p)}\tilde{\mathcal{H}}(\underline{p}) + \int_{\underline{p}}^{p} e^{\frac{\delta+d+\eta-z_X}{z_I-\zeta}(\tilde{p}-p)}\frac{\mu_e}{z_I-\zeta}(1 - \Phi(\exp(\tilde{p})))d\tilde{p},$$

for each $\underline{p}$. Taking the limit $\underline{p} \to -\infty$, and using (18) to replace μ_e with $\delta + d + \eta - z_X$, we arrive at

$$\tilde{\mathcal{H}}(p) = \int_{-\infty}^{p} e^{\frac{\delta+d+\eta-z_X}{z_I-\zeta}(\tilde{p}-p)}\frac{\delta + d + \eta - z_X}{z_I - \zeta}(1 - \Phi(\exp(\tilde{p})))d\tilde{p}. \tag{24}$$

This expression shows $\tilde{\mathcal{H}}(\log y)$ is the complementary cumulative distribution function of a random variable $\mathbf{Y^e}$ defined by a convolution between a Pareto distribution with scale parameter 1 and tail index $(\delta + d + \eta - z_X)/(z_I - \zeta)$ and a distribution with cumulative distribution function (CDF) Φ. That is, $\mathbf{Y}^e$ is expressed as $\mathbf{Y}^e = \mathbf{Y}_1^e\mathbf{Y}_2^e$, where $\mathbf{Y}_1^e \sim \mathit{Pareto}\,(1, (\delta + d + \eta - z_X)/(z_I - \zeta))$ and $\mathbf{Y}_2^e \sim \Phi$.

Notice also that, when Φ is a log-normal distribution, $\tilde{\mathcal{H}}$ is a convolution of a Pareto distribution and a log-normal distribution analyzed in Reed (2001), and more recently, Cao and Luo (2017) and Sager and Timoshenko (2019). Therefore, we offer an alternative micro-foundation of this convolution distribution with an endogenous establishment growth rate, relative to the micro-foundation in Reed (2001) with an exogenous growth rate. Our micro-foundation is also more general because it allows for any distribution of Φ, whereas Reed (2001) only allows Φ to be a log-normal distribution.

Using this explicit solution, we can easily show that when Φ has a thin right tail, for example, when Φ is a (left-truncated) log-normal distribution, and $z_I > \zeta$, $\bar{\mathcal{H}}(p)$ has a Pareto tail with the index given by

$$\lambda^e \equiv \frac{d + \eta - z_X + \delta}{z_I - \zeta}. \tag{25}$$

When $z_I < \zeta$, the distribution has a thin tail.

For the distribution of the number of establishments per firm, (21) becomes

$$0 = -(z_X + \delta + d + \eta)\bar{\mathcal{M}}(1) + 2\delta\bar{\mathcal{M}}(2) + \mu_e \tag{26a}$$

and

$$0 = -(n(z_X + \delta) + d + \eta)\bar{\mathcal{M}}(n) + (n+1)\delta\bar{\mathcal{M}}(n+1) + (n-1)z_X\bar{\mathcal{M}}(n-1) \tag{26b}$$

for $n > 1$. Luttmer (2011) provides a closed-form solution for $\{\bar{\mathcal{M}}(n)\}_{n=1}^{\infty}$:

$$\bar{\mathcal{M}}(n) = \frac{1}{n}\frac{\mu_e}{z_X}\sum_{k=0}^{\infty}\frac{1}{\upsilon_{n+k}}\left(\prod_{m=n}^{n+k}\upsilon_m\right)\prod_{m=1}^{n+k}\frac{z_X\upsilon_m}{\delta}, \tag{27}$$

where the sequence $\{\upsilon_n\}_{n=0}^{\infty}$ is defined recursively by $\upsilon_0 = 0$ and

$$\frac{1}{\upsilon_{n+1}} = 1 - \frac{z_X\upsilon_n}{\delta} + \frac{\eta + d + z_X n}{\delta n}.$$

The distribution of establishment number is given by a discrete random variable $\mathbf{X}$ with pdf

$$\Pr(\mathbf{X} = n) = \frac{\bar{\mathcal{M}}(n)}{\sum_{n'}\bar{\mathcal{M}}(n')}. \tag{28}$$

Because we normalize the measure of n-establishment firms by the total number of establishments $N(t)$, $\sum_n n\bar{\mathcal{M}}(n) = 1$, and thus $\sum_n \bar{\mathcal{M}}(n) < 1$.

In Appendix B.2, we show that when $z_X > \delta$, this distribution has a Pareto tail with the tail index given by

$$\frac{\eta + d}{z_X - \delta}. \tag{29}$$

The following proposition summarizes the last two results. The proof of this proposition is given in Appendix B.2 using a Karamata theorem from the literature on regular varying sequences and functions (Bingham et al. (1987)).[34]

Proposition 2 *On a stationary BGP with $z_I > \zeta$ and $z_X > \delta$ and the distribution of entrant sizes Φ has a thin tail, the stationary distribution of establishment sizes (across establishments) and the stationary distribution of the number of establishments per firm (across firms) have Pareto right tails and the tail indexes are given by* (25) *and* (29)*, respectively.*

Now, we analyze the distribution of firm size. In a special case where the initial draw satisfies $\int \hat{q}d\Phi(\hat{q}) = 1$, (17) implies $z_I = \zeta$ holds; that is, the growth rate of each establishment's product quality is the same as the growth rate of average quality. Because the establishment-size distribution is identical to the distribution of $\hat{q} = q/Q$, and in this case, $\hat{q}$ remains constant over an establishment's lifetime, the intensive-margin distribution is given by the distribution of initial size Φ. In addition, all establishments in a firm have the same size, so the random variable for firm size

[34] Notice this result on the distribution of establishment number is stronger than the one in Luttmer (2011). In particular, he shows that for any $\xi > \frac{\eta+d}{z_X-\delta}$, $\lim_{n\to\infty} n^{\xi}\Pr\{\mathbf{X} > n\} = \infty$ and for any $\xi < \frac{\eta+d}{z_X-\delta}$, $\lim_{n\to\infty} n^{\xi}\Pr\{\mathbf{X} > n\} = 0$, whereas we show

$$\lim_{n\to\infty} n^{\frac{\eta+d}{z_X-\delta}}\Pr(\mathbf{X} > n) = a$$

for some $a > 0$. The latter limit implies the previous two limits but not vice versa.

$\mathbf{Z}$ can be written as $\mathbf{Z} = \mathbf{XY}$, where $\mathbf{X}$ is the number of establishments and $\mathbf{Y}$ is establishment size in a firm. In the cross section, $\mathbf{X}$ and $\mathbf{Y}$ are independent and the cdf of $\mathbf{Y}$ is given by Φ. The pdf for $\mathbf{X}$ is given by (28).

Therefore, the fraction of firms with size $\mathbf{Z}(t) \geq \hat{\mathbf{z}}$, denoted by $\mathbf{M}(\hat{\mathbf{z}})$, can be computed as[35]

$$\mathbf{M}(\hat{\mathbf{z}}) = \frac{\sum_n \bar{\mathcal{M}}(n)(1 - \Phi(\hat{\mathbf{z}}/n))}{\sum_n \bar{\mathcal{M}}(n)}.$$

To determine the tail index of $\mathbf{M}(\cdot)$, we consider the Laplace transformation:[36]

$$\varphi(s) = \int_0^\infty \hat{\mathbf{z}}^s \left(-d\mathbf{M}(\hat{\mathbf{z}})\right). \tag{30}$$

Using the expression for $\mathbf{M}$ above, we show in Appendix B.2 that

$$\varphi(s) = \left\{ \int_0^\infty \hat{\mathbf{z}}^s d\Phi(\hat{\mathbf{z}}) \right\} \left\{ \frac{\sum_n \bar{\mathcal{M}}(n) n^s}{\sum_n \bar{\mathcal{M}}(n)} \right\}.$$

Assuming that the entry distribution Φ has a thin tail and using the characterization in Proposition 2, we show in Appendix B.2 that

$$\varphi(s) \propto \frac{1}{\frac{\eta+d}{z_X - \delta} - s}$$

as $s \uparrow \frac{\eta+d}{z_X-\delta}$. Therefore, by the Tauberian theorem in Mimica (2016, Corollary 1.3), $\mathbf{M}$ has a Pareto tail with the tail index given by (29).

The following proposition summarizes the result.

Proposition 3 *On a stationary BGP with $z_I = \zeta$ and $z_X > \delta$, and when the distribution of entry sizes Φ has a thin tail, the firm-size distribution has a Pareto tail with the tail index equal to the tail index of the distribution of the number of establishments per firm given by* (29)*.*

When the distribution of entry sizes Φ has a Pareto tail, we can show the tail index of the firm size distribution is the minimum of the tail index of Φ and the tail index given by (29).

Now we consider the case in which $z_I \neq \zeta$, i.e. $\int \hat{q} d\Phi(\hat{q}) \neq 1$. This case is more challenging because the dynamics of firm size are driven both by the dynamics of the establishment number and of the dynamics of relative establishment size. This fact implies that when we write firm size as a product of the number of establishments and average establishment size, $\mathbf{Z} = \mathbf{XY}$, in the cross section, $\mathbf{X}$ and $\mathbf{Y}$ are correlated, instead of being independent when $z_I = \zeta$. For example, when $z_X > \delta$ and $z_I > \zeta$, over time surviving firms, on average, have both a higher number of establishments and larger establishments.

To tackle this case, we use the system of differential equations (23). The system of differential

[35] Another way to obtain this result is to notice that $\mathcal{M}(n, \hat{q}) = \bar{\mathcal{M}}(n)(1 - \Phi(\hat{q}))$ solves (23).

[36] After a change of variable $\hat{\mathbf{z}} = \exp(p)$, the transformation can be re-written in its more familiar form: $\int_{-\infty}^{\infty} e^{sp} \left(-d\mathbf{M}(\exp(p))\right)$.

equations (23) for $\mathcal{M}(n, \hat{q})$ simplifies to

$$(z_I - \zeta)\hat{q}\frac{d\mathcal{M}(1, \hat{q})}{d\hat{q}} = -(z_X + \delta + d + \eta)\mathcal{M}(1, \hat{q}) + 2\delta\mathcal{M}(2, \hat{q}) + \mu_e(1 - \Phi(\hat{q}))$$

(for $n = 1$) and

$$(z_I - \zeta)\hat{q}\frac{d\mathcal{M}(n, \hat{q})}{d\hat{q}} = -(n(z_X + \delta) + d + \eta)\mathcal{M}(n, \hat{q}) + (n+1)\delta\mathcal{M}(n+1, \hat{q}) + (n-1)z_X\mathcal{M}(n-1, \hat{q})$$

for $n > 1$. Multiplying both sides of these equations by $\hat{q}^{s-1}$ and integrating by parts from 0 to ∞,

$$\int_0^\infty \hat{q}^{s-1}\mathcal{M}(n, \hat{q})d\hat{q} = -\frac{1}{s}\int_0^\infty \hat{q}^s d\mathcal{M}(n, \hat{q}),$$

we obtain:

$$-(z_I - \zeta)s\hat{\varphi}(1, s) = -(z_X + \delta + d + \eta)\hat{\varphi}(1, s) - 2\delta\hat{\varphi}(2, s) + \int_0^\infty \hat{q}^{s-1}\mu_e(1 - \Phi(\hat{q}))$$

(for $n = 1$) and

$$-(z_I - \zeta)s\hat{\varphi}(n, s) = -(n(z_X + \delta) + d + \eta)\hat{\varphi}(n, s) + (n+1)\delta\hat{\varphi}(n+1, s) + (n-1)z_X\hat{\varphi}(n-1, s)$$

for $n > 1$, where

$$\hat{\varphi}(n, s) \equiv \int_0^\infty \hat{q}^s(-d\mathcal{M}(n, \hat{q})).$$

For each $s \geq 0$, the equations form a system of difference equations and allow us to solve for $\hat{\varphi}(n, s)$ for all $n \geq 1$ using the closed-form solution from Luttmer (2011) (with η being replaced by $\eta - (z_I - \zeta)s$). We show in Appendix B.2 that

$$\hat{\varphi}(n, s) \propto_{n\to\infty} n^{-\frac{d+\eta-(z_I-\zeta)s}{z_X-\delta}-1}.$$

Now, with the solution for $\hat{\varphi}(n, s)$, we can calculate the Laplace transform (30) as follows:

$$\varphi(s) = \frac{1}{\sum_n \bar{\mathcal{M}}(n)}\sum_n n^s \int_0^\infty (\hat{z}/n)^s(-d\mathcal{M}(n, \hat{z}/n)) = \frac{\sum_n n^s\hat{\varphi}(n, s)}{\sum_n \bar{\mathcal{M}}(n)}.$$

Using the asymptotic property of $\hat{\varphi}(n, s)$ above, we show in Appendix B.2 that if $z_X > \delta$ and $z_X - \delta + z_I - \zeta > 0$, $\varphi(s)$ is finite up to s^* determined by

$$\frac{d + \eta - (z_I - \zeta)s}{z_X - \delta} = s,$$

or, equivalently,

$$\lambda^{ne} \equiv \frac{\eta + d}{z_X - \delta + z_I - \zeta}. \tag{31}$$

In addition, we can show

$$\varphi(s) \propto_{s \uparrow s^*} \frac{1}{s^* - s}.$$

Therefore, by the Tauberian theorem in Mimica (2016, Corollary 1.3), $\mathbf{M}$ has a Pareto tail with the tail index given by s^*. We summarize these derivations in the following proposition. The detailed proof is given in Appendix B.2.

Proposition 4 *On a stationary BGP with $z_X > \delta$ and $z_X - \delta + z_I - \zeta > 0$, the firm-size distribution has a Pareto tail with the tail index given by (31). If, in addition, $z_I > \zeta$, all three distributions for establishment size, establishment number, and firm size have a Pareto tails and*

$$\frac{1}{\lambda^f} = \frac{1}{\lambda^{ne}} + \frac{1}{\lambda^e} - \frac{1}{\lambda^{ne}\lambda^e} \tag{32}$$

where λ^e, λ^{ne}, and $\lambda^f > 1$ are respectively the Pareto tail indexes for the distributions for establishment size, establishment number, and firm size.

Because $\lambda^e, \lambda^{ne} > 1$, equation (32) shows $\lambda^f > \lambda^e, \lambda^{ne}$; that is, the tail of the firm-size distribution is strictly fatter than the tails of either the establishment-size distribution or the establishment-number distribution. The formula can also be used to calculate the tail index of the firm-size distribution from the tail indexes for the establishment size and establishment number distributions.[37]

5.3 Discussion for one-type economy

The model features thick tails for all three distributions of interest (total number of workers in the firm, number of workers per establishment within the firm, and number of establishments within the firm). The main mechanism that generates these tails is firm growth with random exit. As a firm grows by adding establishments and adding workers to those establishments, a Poisson arrival of exit shocks (δ and d) occurs that prevents the firm from continuing to grow. Therefore, the relatively small mass of lucky firms that do not receive an exit shock will continue to grow. The luckiest firms will eventually grow to be very large relative to the average firm and create a thick upper tail of the firm-size distribution(s). For the one-type characterization in Section 5.2, the expressions for the Pareto tails reflect these forces.

From Proposition 2 and equation (29), we see the distribution over the number of establishments per firm has a Pareto tail equal to $(\eta + d)/(z_X - \delta)$. This expression relates the aggregate inflow of new establishments and outflow of firms through random exit to the net growth rate for establishments ($z_X - \delta$). The distribution's tail is fatter when additional establishments accumulate faster,

[37] Equation (32) approximates the relationships of the estimated tail indexes in the data very well. For example, for 1995, with the tail index of the establishment number distribution of 1.25 in Table 1 and the tail index of the establishment size distribution of 1.40 from the estimation in Appendix D.3 (which is conceptually different from the tail index for the intensive margin in Table 1), equation (32) implies a tail index of 1.06 for firm-size distribution, which lies between the estimates 0.99 and 1.10 for firm size distribution in Table 1. Similarly, for 2015, with the tails parameters of 1.44 and 1.21, equation (32) implies a tail index of 1.05 for firm size distribution, which also lies between the estimates of 0.99 and 1.17 in Table 1.

given random exits induced by establishment and firm shocks δ and d. Using the decomposition of the final-output growth rate in equations (17) and (18), we can rewrite the Pareto tail for the distribution of establishment number as

$$1 + \frac{\mu_e}{z_X - \delta}.$$

This expression shows the upper tail of the establishment-number distribution is fatter when less firm entry occurs and incumbent firms' net growth is faster. Both elements suggest that tail fatness is related to selection on random exits: with fewer firms due to less entry, a fat upper tail only emerges if firms grow fast enough during their finite lifetimes.

Similarly, from equation (25), we see the distribution of establishment size has a Pareto tail index equal to $[(\eta + d)/(z_I - \zeta)] - [(z_X - \delta)/(z_I - \zeta)]$. This expression is similar to that for the establishment-number distribution. The first term describes a fat upper tail as arising from low churn in terms of establishment inflow and firm death relative to high net growth in establishment size $(z_I - \zeta)$, where growth in intensive-margin investments to firm quality is large relative to aggregate quality growth ζ due to the effect of entrants. The second term is simply an adjustment for the net growth in establishment number. Again using equations (17) and (18), the Pareto tail index can be rewritten as

$$\frac{1}{1 - \int_{\mathcal{N}} \hat{q} d\Phi(\hat{q})}.$$

This version of the expression makes clear that the distribution exhibits a fatter tail when the average entering establishment size is small relative to the average incumbent establishment, so that $\int \hat{q} d\Phi(\hat{q})$ is far less than 1. In this case, the establishment must grow over time to catch up to the average incumbent establishment, during which time random exits pare down the mass of growing establishments. The more that entrants need to catch up to the average incumbent, the more inequality we observe in the distribution of establishment size.

Finally, we can study the characterization of the overall firm-size distribution. From Proposition 4 and equation (32), the Pareto tail of the firm-size distribution is equal to $(\eta + d)/(z_X - \delta + z_I - \zeta)$, which tells us a fat upper tail emerges when the firm grows fast – on either margin – relative to the churn of establishment inflows and firm deaths $(\eta + d)$. This expression can also be rewritten using equations (17) and (18):

$$\frac{1}{1 - \frac{\mu_e}{z_X - \delta + \mu_e} \int_{\mathcal{N}} \hat{q} d\Phi(\hat{q})},$$

so that the distribution exhibits a fatter upper tail when entrants are small relative to the average incumbent (an effect that derives from the establishment-size distribution) and when the entry rate of new firms is low relative to net growth in establishment creation (an effect that derives from the establishment-number distribution).

We can combine these observations with the comparative statics in Proposition 1 to obtain comparative statics of the tail indexes. For example, the following proposition shows how the tail indexes vary with changes in the entry cost parameter ϕ.

Proposition 5 *An increase in the entry cost, ϕ, makes the tail of the establishment-number distribution and the firm-size distribution fatter, i.e., more inequality at the top, whereas the tail of the establishment-size distribution remains unchanged.*

5.4 Distributions for two-type economy

In Appendix C.1, we show the one-type model, when estimated, produces too few single-establishment firms (even though it matches very well the Pareto-tail index of the distribution of establishment numbers and the distribution of establishment sizes in the data). Therefore, we need at least two types to generate salient features in the data.

Consider an economy with two types: $\tau \in \{L, H\}$. To simplify the derivations of stationary distributions, we assume $0 = \lambda_{LH} < \lambda_{HL}$ (as we do in Section 6); that is, the H-type can transition to becoming an L type, but not vice versa.

Using the results for two types in Luttmer (2011) and the derivations in Subsection 5.2, we can show that, under some parameter restrictions, the stationary distribution of the number of establishments per firm hasthe Pareto tail with the tail index given by:

$$\min\left\{\frac{\eta + \lambda_{HL} + d_H}{[z_X^H - \delta_H]_+}, \frac{\eta + d_L}{[z_X^L - \delta_L]_+}\right\}, \tag{33}$$

where $[x]_+ \equiv max(x, 0)$. The formula corresponds to (29) in the case of a single type.

Similarly, using the procedure from Gabaix et al. (2016) and Cao and Luo (2017), the Pareto-tail index of the distribution of establishment size is given by

$$\min\left\{\frac{\eta + \delta_H + \lambda_{HL} + d_H - z_X^H}{[z_I^H - \zeta]_+}, \frac{\eta + \delta_L + d_L - z_X^L}{[z_I^L - \zeta]_+}\right\}. \tag{34}$$

Lastly, assuming $z_X^L \leq \delta_L$, using a Laplace transform as in Section 5.2, we can show the Pareto-tail index of the firm-size distribution is given by

$$\frac{\eta + \lambda_{HL} + d_H}{[z_X^H - \delta_H + z_I^H - \zeta]_+}. \tag{35}$$

6 Quantitative analysis

In this section, we estimate our model to quantitatively analyze the pattern of firm growth over the 1995-2014 period. We confirm the model is able to match the firm-size, establishment-number, and establishment-size distributions well. We estimate the model for both 1995 and 2014 data and show how the model informs us about fundamental economic forces that changed with average firm-size growth over these years.

6.1 Model estimation

First, we describe our two-step estimation procedure. Then, we present the estimates and associated quantitative results.

6.1.1 Computing equilibrium

Because we can solve for g from (10) as a function of parameters (exogenous growth), we can obtain r directly from the representative households' Euler equation (1). After solving the HJB equations and finding $\bar{w}$ that satisfies the free-entry condition, we can use (15) and (16) to obtain μ_e and s_τ. Given these parameter values, η and M_τ can be solved from (13) and (21).

6.1.2 Overview of the estimation procedure

Although the analytical characterization of the model with one type provided us useful insights, Appendix C.1 shows at least two types are needed to match both the extensive- and intensive-margin distributions. As such, we estimate a simple version of the model with two types, denoted $\tau \in \{L, H\}$. H-type firms turn out to have a lower cost of external investment, and expand their number of establishments faster. We assume H-type ("high type") firms transition to L-type ("low type") firms at the rate $\lambda_{HL} > 0$, whereas L-type firms do not transition to H-type firms, that is, $\lambda_{LH} = 0$. Thus, the L-type is the absorbing state. These assumptions are similar to those in Luttmer (2011). These assumptions allow us to easily obtain closed-form solutions for the tail indexes of the distributions of establishment number and establishment size. We use these closed-form solutions to estimate the model. Under these assumptions, Luttmer (2011) also provides analytical solutions for the distribution of establishment numbers, which can be used to verify the accuracy of numerical solutions.

We estimate the model in two steps. In Step 1, we estimate $(z_X^H, z_X^L, \lambda_{HL}, \mu_e, m_H, m_L)$ using moments related to the number of establishments per firm (Step 1a), and then we estimate $(z_I^H, z_I^L, \Phi(\cdot))$ using moments related to the number of employees per establishment (Step 1b). In Step 2, we assume functional forms for the cost functions $h_X^\tau(\cdot)$, $h_I^\tau(\cdot)$ and estimate the parameters of these functions using the estimates from Step 1.

Step 1a (Number of establishments per firm): In this step, we choose

$$(z_X^H, z_X^L, \lambda_{HL}, \mu_e, m_H, m_L)$$

parameters to target (i) percentiles of the distribution over the number of establishments per firm, (ii) the slope of the upper tail of the distribution, and (iii) the growth rate of the number of establishments $\eta \approx 1\%$. The empirical moments are described in Appendix D.2.

With two types, (13) becomes

$$\eta = z_X^H - \delta_H - d_H + \mu_e \frac{m_H}{M_H} - \lambda_{HL}$$
$$\eta = z_X^L - \delta_L - d_L + \mu_e \frac{m_L}{M_L} + \lambda_{HL} \frac{M_H}{M_L}.$$

Together with $M_H + M_L = 1$, the last equality gives us a unique solution for M_H, M_L. In particular,

$$M_L = \frac{-a_1 - \sqrt{a_1^2 - 4a_0 a_2}}{2a_0},$$

where

$$\begin{aligned} a_0 &= (z_X^H - \delta_H - d_H) - (z_X^L - \delta_L - d_L) \\ a_1 &= -(\mu_e(m_H + m_L) + \lambda_{HL}) + ((z_X^L - \delta_L - d_L) - (z_X^H - \delta_H - d_H)) \\ a_2 &= \mu_e m_L + \lambda_{HL}, \end{aligned}$$

and $M_H = 1 - M_L$. We then obtain η from either equation.

Having η, we can use equation (33) to calculate the Pareto-tail index of the distribution of establishment number. We also use (21) to compute the whole distribution of establishment numbers including the fraction of single-establishment firms and several quantiles beyond the top 99%. Using these model moments, the estimation minimizes a weighted squared sum between the model and empirical moments.

Step 1b (Establishment size): In this step, we assume the entry-size distribution Φ follows a log-normal distribution with mean ϱ and variance ς^2:

$$\Phi \sim \exp(\mathcal{N}(\varrho, \varsigma^2)).$$

We choose

$$z_I^H, z_I^L, \varrho, \varsigma$$

and target the distribution of establishment size as well as the average growth rate of establishment size,

$$\zeta = g - \eta.$$

The empirical moments from the establishment size distribution include the Pareto tail index and several quantiles. These moments are computed from the estimated parameterized distribution described in Appendix D.3. The estimation of the parametrized distributions uses publicly-available BLS data.

In the model, the growth rate ζ can be computed from (16). The tail index of the establishment size distribution in the model is given by (34). The whole distribution of establishment size can be computed by solving (22), and several quantiles are included in the list of targeted moments.

Using these model moments, the estimation minimizes a weighted-squared-sum distance between the model and empirical moments as in Step 1a.

Step 2 (Recovering endogenous variables): In this step, we use the estimates from Step 1a and Step 1b, including $z_i^\tau, \tau \in \{H, L\}, i \in \{X, I\}$, to quantify the remaining model outcomes and allocations. To execute this step, we must parameterize the innovation-cost functions. We assume the innovation-cost functions take the form $h_i^\tau(z) = \chi_i^\tau z^\psi$, for $\tau \in \{L, H\}$, $i \in \{X, I\}$, where $\psi > 0$. The first-order condition in (7) implies $\psi \chi_i^\tau (z_i^\tau)^{\psi-1} = v_\tau$, and hence

$$-h_i^\tau(z_i^\tau) + z_i^\tau v_\tau = \left(1 - \frac{1}{\psi}\right) z_i^\tau v_\tau.$$

Substituting this expression in (7) and re-arranging, we arrive at:

$$\underbrace{\begin{bmatrix} A_{11} & -\lambda_{HL} \\ 0 & A_{22} \end{bmatrix}}_{\equiv \boldsymbol{A}} \begin{bmatrix} v_H \\ v_L \end{bmatrix} = \bar{\pi} \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

where

$$A_{11} = r - \left(1 - \frac{1}{\psi}\right)(z_X^H + z_I^H) + \delta_H + d_H + \lambda_{HL}$$

and

$$A_{22} = r - \left(1 - \frac{1}{\psi}\right)(z_X^L + z_I^L) + \delta_L + d_L + \lambda_{HL}.$$

From the estimates in Step 1a and Step 1b, all the elements of matrix $\boldsymbol{A}$ are known, including $r = \rho + \sigma g$ and $g = \eta + \zeta$. Therefore, we can then solve for v_H, v_L as functions of $\bar{\pi}$:

$$\begin{bmatrix} v_H \\ v_L \end{bmatrix} = \bar{\pi} \boldsymbol{A}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Now, combining this result with equations (8) and (9), we obtain

$$\phi = \bar{\pi} \begin{bmatrix} m_H \\ m_L \end{bmatrix}' \boldsymbol{A}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \exp\left(-\varrho + \frac{\varsigma^2}{2}\right).$$

In other words, $\bar{\pi}$ is uniquely determined as a function of ϕ.

Lastly, we choose ϕ to match the empirical investment-to-output ratio. On a BGP, we can decompose $Y(t)$ into $C(t), R(t), E(t)$ and obtain the following expression for the investment-to-output ratio:

$$\frac{R(t) + E(t)}{Y(t)} = \left(\mu_e \phi + \sum_\tau s_\tau \left(h_X^\tau(z_X^\tau) + h_I^\tau(z_I^\tau)\right)\right) \left(\frac{\bar{w}(t)}{1 - \beta}\right)^{\frac{1-\beta}{\beta}}. \tag{36}$$

Table 2: Parameter Values and Targets

Concept	Parameter	Value	Target/Source
Parameters Set in Advance			
Elasticity of Demand	β	$1-(1/1.10)$	10% Markup
Intertemporal Elasticity	σ	1	Log utility
Discount Rate	ρ	0.01	Standard value
Population Growth Rate	γ	0.01	Census Bureau
Firm Exit Rates	d_L, d_H	0.4%, 0%	BLS
Establishment Exit Rates (1995)	δ_L, δ_H	12%, 12%	BLS
Establishment Exit Rates (2014)	δ_L, δ_H	10%, 10%	BLS
Innovation Cost	ψ	2	Quadratic baseline
Estimated Parameter Targets			
Entry Cost	ϕ		$\frac{Investment}{Output} = 10\%$, NIPA
Extensive Margin Costs	χ_X^H, χ_X^L		Establishment number distribution
Intensive Margin Costs	χ_X^H, χ_X^L		Establishment size distribution
Growth Types	λ_{HL}, m_H		Establishment number/size right tail
Entrant Size	ϱ, ς		Establishment size distribution

The derivation of this formula is presented in Appendix A.2.

6.1.3 Parameter values and estimates

A set of parameters are assigned in advance of estimation, whereas the remaining parameters are estimated to match empirical moments of the establishment number and establishment size distributions. Table 2 summarizes parameter estimates and targets.

The unit of time is set as a year. For preferences, for simplicity, we assume log utility and an effective discount rate of $\rho = 0.01$. We choose the elasticity of demand for the final-good producer to $\beta = 0.091$, which is consistent with a markup of 10% (Basu and Fernald (1995)). We assume quadratic innovation costs by setting the elasticity to $\psi = 2$ as a baseline assumption. With respect to firm and establishment exit rates, we set $d_H = 0\%$ and $d_L = 0.4\%$ at the firm level and $\delta_H = \delta_L = 12\%$ in 1995 and $\delta_H = \delta_L = 10\%$ in 2015. These values amount to around a 3% quarterly exit rate for establishments and a 0.1% quarterly (exogenous) exit rate for firms. Lastly, we set the population growth rate to the post-1960 average of $\gamma = 0.01$ and assume purely exogenous growth $\alpha = 1$ as a baseline assumption.

The remaining parameters are estimated to match empirical moments. We target $\eta = 0.01$ and $\zeta = 0.02$ to imply a growth rate of final output of 3% for 1995 and $\eta = 0.01$ and $\zeta = 0.01$, implying a final output growth rate of 2% for 2015. With $\rho = 0.01$ and log utility, this calibration implies $r = 0.04$ in (1) for 1995 and $r = 0.03$ for 2014, which are within the range of values that are standard in the literature. We choose an entry cost, ϕ, to match total investment as 10% of total output.

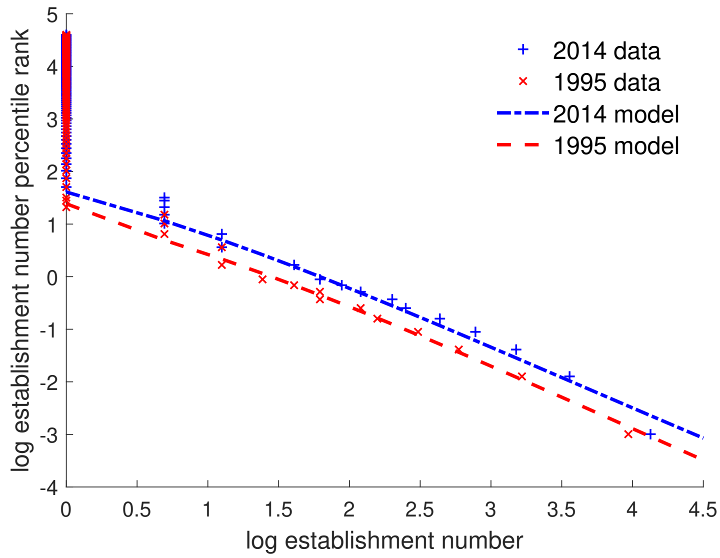


Figure 6: Distribution of number of establishments per firm, Data and Model

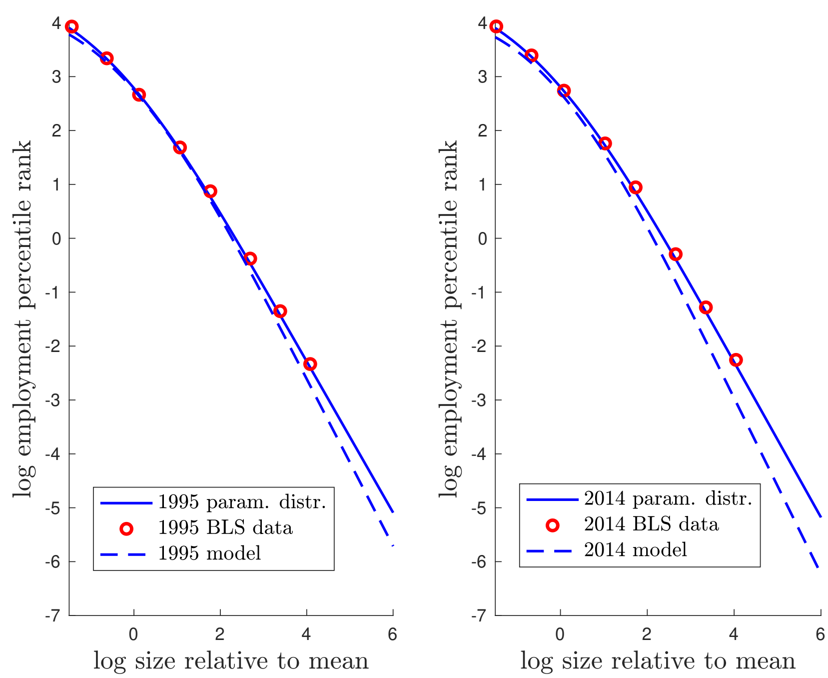


Figure 7: Distribution of number of employees per establishment, Data and Model

Table 3: Moment Fitness

Moments	Data	Model	Data	Model
	1995		2014	
Step 1a: Establishment Number Distribution				
Establishment growth rate (η)	0.010	0.010	0.010	0.010
Pareto tail index	1.25	1.25	1.21	1.21
Step 1b: Establishment Size Distribution				
Average growth rate (ζ)	0.021	0.021	0.013	0.014
Pareto tail index	1.40	1.56	1.44	1.65

Notes: Estimates using Least-Square Minimization.

We follow the two-step procedure described in Section 6.1.2 to estimate $(\chi_L^X, \chi_H^X, \mu_E, \lambda_{HL}, m_H)$ and $(\chi_L^I, \chi_H^I, \varrho, \varsigma)$. Finally, we estimate parameters twice—once for 1995 moments of the establishment size and number distributions, and once for 2014 empirical moments.

Figure 6 shows the model distributions of the number of establishments per firm match the empirical distributions from the 1995 and 2014 data very well. Furthermore, Figure 7 shows the model distributions of establishment size closely match the parametrized empirical distributions for 1995 and 2014 data described in Appendix D.3 (blue solid line), as well as publicly available BLS tabulations of establishment sizes (red circles). The success of the model along these dimensions has to do with the existence of fat tails in the data. The model generates fat-tailed distributions endogenously, and therefore the estimation procedure is selecting parameters to match the general slopes of the Pareto tails in the data. Table 3 shows the model distribution closely matches the remaining empirical targets including Pareto-tail estimates computed from the establishment size and number distributions.

6.2 Quantitative results

In this section, we compare the results of model estimation for 1995 with that in 2014. Differences in model outcomes over time inform us about the underlying economic mechanisms that generate the observed changes in the distributions over the number of establishments and average establishment size. We view this exercise as a growth decomposition of the data through the lens of our theory.

To begin the comparison between 1995 and 2014 outcomes, Table 4 shows the H-type firm's extensive-margin investment rate increased from 32.81% to 51.20% as its investment-cost coefficients decreased from 0.7830 to 0.6366. This finding is not too surprising, because the theory suggests the growth in the concentration of firms in the upper tail of the establishment number distribution is partially driven by the increased extensive-margin investment. Yet, the model simultaneously uncovers that the high-type firm's intensive-margin investment rate was zero in both 1995 and 2014, which is driven by very high investment costs in the model. From the perspective of the

Table 4: Parameter Estimates and Model Outcomes, 1995 versus 2014

Parameter	Description	Value (1995)	Value (2014)
Innovation Investments			
z_X^H	H-type external innovation	0.3281	0.5120
z_X^L	L-type external innovation	0.0019	0.0002
z_I^H	H-type internal innovation	0.0000	0.0000
z_I^L	L-type internal innovation	0.1058	0.0822
Innovation Costs			
χ_X^H	H-type external innovation cost	0.7830	0.6366
χ_X^L	L-type external innovation cost	94.2434	941.8146
χ_I^H	H-type internal innovation cost	∞	∞
χ_I^L	L-type internal innovation cost	1.6615	2.4823
Firm Entry			
μ_e	Entry rate	0.0980	0.0739
ϕ	Entry fixed cost	0.1452	0.1855
$\int \hat{q} d\Phi(\hat{q})$	Entrant size relative to mean	0.4031	0.4317
ϱ	Mean of $\Phi(\cdot)$	-1.7638	-1.5964
ς	Standard deviation of $\Phi(\cdot)$	1.3079	1.2300
Firm Types			
λ_{HL}	H to L transition rate	0.2523	0.4900
m_H	Fraction of H-type at entry	0.0523	0.0878
m_L	Fraction of L-type at entry	0.9477	0.9122

Notes: Estimates using Least-Square Minimization

model, if an entering firm draws the H-type, it has an incentive to invest entirely on the extensive margin to take advantage of low R&D costs. The firm invests in intensive margin innovation after it receives the transition shock λ_{HL}. The L-type firms invest little along the extensive margin (0.19% in 1995) and more than H-type firms along the intensive margin (10.58% in 1995). Furthermore, as the establishment-size distribution from the data displays a thinner upper tail over time and the theory predicts that the establishment-size distribution's upper tail is partially driven by intensive-margin investment rates, we see that L-type firms invest less in 2014 (8.22%) than in 1995. Finally, the model shows L-type firms decreased their extensive margin investment rate from 0.19% in 1995 to 0.02% in 2014.

The model finds that over time more entrants are high-type firms (5.23% in 1995 and 8.78% in 2014), yet the average duration of being a high-type shortened from around four years (= 1/0.2523) in 1995 to around two years (= 1/0.4900) in 2014. Although this shortens the periods of high growth along the extensive margin, high-type firms also have an increasing incentive to heavily innovate before becoming low-type firms.

With respect to entry, the estimation provides estimates of μ_e, which is the entry rate of firms relative to the number of establishments. To recover the firm-entry rate (relative to the number of firms), we need to divide μ_e by the average number of establishments per firm. As a result, the model recovers a declining firm-entry rate of 7.55% in 1995 to 5.08% in 2014. This decline is accompanied by an increase in the estimated entry cost. A declining entry rate is consistent with recent empirical evidence in Decker et al. (2014).[38] The model also recovers an increase in the average initial size of entering establishments relative to the average incumbent establishments, from 40.31% in 1995 to 43.17% in 2014. This result is consistent with theory, which tells us that an increase in the size of entrants, $\int \hat{q} d\Phi(\hat{q})$, leads to a thinner upper tail of the establishment-size distribution.

With the estimated model, we conduct a counterfactual experiment. We compute the contribution of each parameter to the changes in the entry rate and in the number of establishments per firm, over the period 1995-2014. To do so, we first consider all parameters that were estimated in the year 1995, and then incrementally change each parameter to the estimate from the year 2014. Table 5 shows, using counterfactual decompositions, how the entry rate and the average number of establishments per firm are affected by the changes in parameters from 1995 to 2014. The table, starting from the top row marked "Decomposition," begins from the 1995 estimated parameters and "turns on" the 2014 parameters as we go down each row. At the end of the final row, all parameters are switched to the ones in 2014, and therefore the sum of all rows is equal to the total change.

The first column conducts this exercise for the entry rate. The entry rate decreases by 2.41% in total. This total turns out to be primarily driven by changes in the fixed entry cost and the decline in the establishment exit rate. The second column is the decomposition for the number of establishments per firm (the extensive margin), which is the main focus of this paper. The average number of establishments per firm increases by 12.14%. Our decomposition reveals that this total is primarily driven by the changes in the external innovation cost and the decline in the exit rate.

7 Conclusion

In this paper, we decomposed firm growth into two margins: an extensive margin of building new establishments and an intensive margin of adding workers to existing establishments. We documented the patterns of extensive- and intensive-margin firm growth in the U.S. from 1991-2014 and found that U.S. growth is predominantly generated by the addition of new establishments in very large firms. We developed a model of firm growth that incorporates both the extensive and intensive margins as separate types of firm innovation and showed the model can generate a fat tail of large firms, both in terms of the number of establishments and the number of workers.

We estimate the model parameters for 1995 and 2014, and use the model to interpret the increase in firm sizes that we observe in the data as reflecting fundamental economic changes.[39]

[38] The level of the entry rates in the model are lower than the ones in the data but the magnitude of the decline is similar. We can bring the model entry rates closer to the data by allowing for type-dependent exit rates.

[39] The general version of the model is rich and flexible but for tractability we focus on a two-type version in the

Table 5: BGP Changes and Decomposition

		Δ in entry rate (%)	$\%\Delta$ in #estab/firm
Aggregate 1995-2014		-2.47	12.14
Decomposition:			
type fraction and persistence	(m_H, λ_{LH})	2.63	-19.05
entrant quality distribution	(ϱ, ς)	0.96	-0.68
fixed entry cost	(ϕ)	-6.20	3.71
external innovation cost	(χ_X^H, χ_X^L)	-0.51	18.54
internal innovation cost	(χ_I^H, χ_I^L)	6.41	-4.87
establishment exit rates	(δ_H, δ_L)	-4.18	11.05
growth rate	(g)‘	-1.57	7.40

We find that the cost for external innovation declined for firms that are actively expanding with new establishments, whereas the internal innovation cost has increased. The entry cost is higher now, whereas the average quality at entry has improved. Upon entry, growth through external innovation became more common, but the duration of such a high-growth period is shorter now. A decomposition exercise revealed the largest contributors to the recent dominance of large firms with many establishments are the decline in the external innovation cost and the decline in the exit rate.

An important future research agenda is to explain why these changes occurred during this time period. Various anecdotes indicate finding new locations for stores and restaurants became easier due to increasing availability of data. This may have contributed to the lower cost of external innovation. Faster information flows better enables a new business to succeed early, but also allows business models to be imitated more easily, creating a faster obsolescence and hampering the ability to grow quickly through the extensive margin. This type of anecdote is consistent with the shorter duration of the high-growth period. The recent literature also points out increases in regulations and entry barriers that contributed to the decline of dynamism in the U.S. economy. These changes are consistent with the estimated increase in the entry cost in our model.

estimation. Having more types would allow us to match more firm-establishment facts including the growth patterns documented in Appendix E.2

References

Abowd, J., B. Stephens, L. Vilhuber, F. Andersson, K. McKinney, M. Roemer, and S. Woodcock (2009). The LEHD infrastructure files and the creation of the quarterly workforce indicators. In T. Dunne, J. B. Jensen, and M. J. Roberts (Eds.), *Producer Dynamics: New Evidence from Micro Data*, Volume 68, Chicago, IL, pp. 149–230. Studies in Income and Wealth: University of Chicago Press.

Acemoglu, D. and D. Cao (2015). Innovation by Entrants and Incumbents. *Journal of Economic Theory 157*, 255–294.

Aghion, P., A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li (2019). A Theory of Falling Growth and Rising Rents. NBER Working Paper 26448.

Akcigit, U. and W. R. Kerr (2018). Growth through Heterogeneous Innovations. *Journal of Political Economy 126*(4), 1374–1443.

Akicigit, U. and S. T. Ates (2019). What Happened to the U.S. Business Dynamism? NBER Working Paper 25756.

Argente, D., M. Lee, and S. Moreira (2019). How Do Firms Grow? The Life Cycle of Products Matters. Working paper.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2017). The Fall of the Labor Share and the Rise of Superstar Firms. NBER Working Paper 23396.

Axtell, R. and O. Guerrero (2019). *Dynamics of Firms: Data, Theories, and Models*. MIT Press.

Axtell, R. L. (2001). Zipf Distribution of U.S. Firm Sizes. *Science 293*(5536), 1818–1820.

Beare, B. and A. Toda (2017). Geometrically Stopped Markovian Random Growth Processes and Pareto Tails. USCD Working Paper.

Bingham, N., C. Goldie, and J. Teugels (1987). *Regular Variation*. Cambridge University Press.

Cao, D. and W. Luo (2017). Persistent Heterogeneous Returns and Top End Wealth Inequality. *Review of Economic Dynamics 26*, 301 – 326.

Choi, E. J. and J. R. Spletzer (2012). The Declining Average Size of Establishments: Evidence and Explanations. *Monthly Labor Review*, 50–65.

De Loecker, J. and J. Eeckhout (2017). The Rise of Market Power and the Macroeconomic Implications. NBER Working Paper 23687.

Decker, R., J. Haltiwanger, R. S. Jarmin, and J. Miranda (2014). The Role of Entrepreneurship in U.S. Job Creation and Economic Dynamism. *Journal of Economic Perspectives 28*, 3–24.

Dinlersoz, E., S. Kalemli-Ozcan, H. Hyatt, and V. Penciakova (2018). Leverage over the life cycle and implications for firm growth and shock responsiveness. NBER Working Paper 25226.

Gabaix, X. (2009). Power Laws in Economics and Finance. *Annual Review of Economics 1*(1), 255–294.

Gabaix, X. (2011). The Granular Origins of Aggregate Fluctuations. *Econometrica 79*(3), 733–772.

Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll (2016). The Dynamics of Inequality. *Econometrica 84*(6), 2071–2111.

Garca-Macia, D., C.-T. Hsieh, and P. J. Klenow (2019). How Destructive is Innovation? *Econometrica 87*, 1507–1541.

Garetto, S., L. Oldenski, and N. Ramondo (2019). Multinational Expansion in Time and Space. Working paper.

Gutiérrez, G. and T. Philippon (2017). Declining Competition and Investment in the U.S. NBER Working Paper 23583.

Hathaway, I. and R. E. Litan (2014). Declining Business Dynamism in the United States: A Look at States and Metros. Brookings Institute.

Hershbein, B., C. Macaluso, and C. Yeh (2019). Concentration in U.S. Local Labor Markets: Evidence from Vacancy and Employment Data. Working paper.

Holmes, T. J. (2011). The Diffusion of Wal-Mart and Economics of Density. *Econometrica 79*, 253–302.

Hopenhayn, H. and R. Rogerson (1993). Job Turnover and Policy Evaluation: A General Equilibrium Analysis. *Journal of Political Economy 101*(5), 915–938.

Hsieh, C.-T. and E. Rossi-Hansberg (2019). The Industrial Revolution of Services. NBER Working Paper 25968.

Jones (1995). R&D-Based Models of Economic Growth. *Journal of Political Economy*, 759–784.

Kehrig, M. and N. Vincent (2019). Good Dispersion, Bad Dispersion. NBER Working Paper 25923.

Klette, T. J. and S. Kortum (2004). Innovating Firms and Aggregate Innovation. *Journal of Political Economy 112*(5), 986–1018.

Kondo, I. O., L. Lewis, and A. Stella (2019). On the U.S. Firm and Establishment Size Distributions. Working paper.

Lentz, R. and D. T. Mortensen (2008). An Empirical Model of Growth through Product Innovation. *Econometrica 76*, 1317–1373.

Luttmer, E. G. J. (2011). On the Mechanics of Firm Growth. *The Review of Economic Studies 78*(3), 1042–1068.

Mimica, A. (2016). Exponential Decay of Measures and Tauberian Theorems. *Journal of Mathematical Analysis and Applications 440*(1), 266 – 285.

Mukoyama, T. and S. Osotimehin (2019). Barriers to Reallocation and Economic Growth: The Effects of Firing Costs. *American Economic Journal: Macroeconomics 11*, 235–270.

Reed, W. J. (2001). The Pareto, Zipf and Other Power Laws. *Economics Letters 74*(1), 15 – 19.

Rinz, K. (2018). Labor Market Concentration, Earnings Inequality, and Earnings Mobility. CARRA Working Paper 2018-10.

Rossi-Hansberg, E., P.-D. Sarte, and N. Trachter (2018). Diverging Trends in National and Local Concentration. NBER Working Paper 25066.

Sager, E. and O. A. Timoshenko (2019). The Double EMG Distribution and Trade Elasticities. Canadian Journal of Economics Forthcoming.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter (2015). Firming Up Inequality. NBER Working Paper 22953.

Vilhuber, L. (2018). LEHD infrastructure S2014 files in the FSRDC. Technical report. U.S. Census Bureau Center for Economic Studies Working Paper #CES-18-27.

Xi, X. (2016). Multi-establishment Firms, Misallocation and Productivity. Working paper.

Appendix

A Derivations

A.1 Derivation of (16)

From the definition of $Q_\tau(t)$,

$$\frac{\dot{Q}_\tau(t)}{Q_\tau(t)} = -\frac{\dot{N}_\tau(t)}{N_\tau(t)} + \frac{d\int_{\mathcal{N}_\tau(t)} q_j(t)dj/dt}{\int_{\mathcal{N}_\tau(t)} q_j(t)dj}. \tag{37}$$

The first term of the right-hand side is $-\eta$. To compute the second term, consider a discrete time interval $\Delta t > 0$, compute $(\int_{\mathcal{N}_\tau(t+\Delta t)} q_j(t+\Delta t)dj - \int_{\mathcal{N}_\tau(t)} q_j(t)dj)/\Delta t$ and set $\Delta_t \to 0$. Note that the denominator of the second term is equal to $Q(t)N(t)$. Because

$$\begin{aligned}
&\int_{\mathcal{N}_\tau(t+\Delta t)} q_j(t+\Delta t)dj - \int_{\mathcal{N}_\tau(t)} q_j(t)dj \\
&\quad = [z_I^\tau + z_X^\tau]\Delta t Q_\tau(t)N_\tau(t) - (\delta_\tau + d_\tau)\Delta t Q_\tau(t)N_\tau(t) + \mu_e \Delta t m_\tau Q(t)N(t)\int \hat{q}d\Phi(\hat{q}) \\
&\quad\quad - \sum_{\tau'\neq\tau} \lambda_{\tau\tau'}\Delta t Q_\tau(t)N_\tau(t) + \sum_{\tau'\neq\tau} \lambda_{\tau'\tau}\Delta t Q_{\tau'}(t)N_{\tau'}(t) + o(\Delta t),
\end{aligned}$$

where the first term is the additional quality by internal and external innovation, the second term is the lost quality by exit, the third term is the gain from entry, and the fourth and the fifth terms are the loss and gain from the transitions of firm types. (The higher-order terms are omitted as $o(\Delta t)$.) Dividing by Δt and taking $\Delta t \to 0$,

$$\begin{aligned}
&\frac{d\int_{\mathcal{N}_\tau(t)} q_j(t)dj}{dt} \\
&= Q_\tau(t)N_\tau(t)\left(z_I^\tau + z_X^\tau - (\delta_\tau + d_\tau) + \mu_e m_\tau \frac{Q(t)N(t)}{Q_\tau(t)N_\tau(t)}\int \hat{q}d\Phi(\hat{q}) - \sum_{\tau'\neq\tau}\lambda_{\tau\tau'} + \sum_{\tau'\neq\tau}\lambda_{\tau'\tau}\frac{Q_{\tau'}(t)N_{\tau'}(t)}{Q_\tau(t)N_\tau(t)}\right).
\end{aligned}$$

Therefore, (37) can be rewritten as

$$\zeta = -\eta + z_I^\tau + z_X^\tau - (\delta_\tau + d_\tau) + \mu_e \frac{m_\tau}{M_\tau}\frac{Q(t)}{Q_\tau(t)}\int \hat{q}d\Phi(\hat{q}) - \sum_{\tau'\neq\tau}\lambda_{\tau\tau'} + \sum_{\tau'\neq\tau}\lambda_{\tau'\tau}\frac{Q_{\tau'}(t)M_{\tau'}}{Q_\tau(t)M_\tau}$$

Using the definition of s_τ and $g = \eta + \zeta$, we can obtain (16).

A.2 Accounting for Aggregate Resources

In this subsection, we decompose $Y(t)$ into $C(t), R(t), K(t)$ on a BGP.

Plugging the optimal choice for $x_j(t)$, (3), in the aggregate production function (2), we obtain

$$
\begin{aligned}
Y(t) &= \left(\int_{\mathcal{N}(t)} q_j(t)^{\beta} x_j(t)^{1-\beta} dj\right)^{\frac{1}{1-\beta}} \\
&= \left(\int_{\mathcal{N}(t)} q_j(t)^{\beta} \left[(1-\beta)^{\frac{1}{\beta}} \left(\frac{w(t)}{A(t)}\right)^{-\frac{1}{\beta}} Y(t) q_j(t)\right]^{1-\beta} dj\right)^{\frac{1}{1-\beta}} \\
&= (1-\beta)^{\frac{1}{\beta}} \left(\frac{w(t)}{A(t)}\right)^{-\frac{1}{\beta}} Y(t) \left(\int_{\mathcal{N}(t)} q_j(t) dj\right)^{\frac{1}{1-\beta}} \\
&= \left(\frac{w(t)}{(1-\beta)A(t)}\right)^{-\frac{1}{\beta}} Y(t) \left(N(t)Q(t)\right)^{\frac{1}{1-\beta}},
\end{aligned}
$$

where the last equality uses the definition of $Q(t)$.

Simplifying $Y(t)$ from both sides, we arrive at:

$$
1 = \left(\frac{w(t)}{(1-\beta)A(t)}\right)^{-\frac{1}{\beta}} \left(N(t)Q(t)\right)^{\frac{1}{1-\beta}}.
$$

Combining this identity with the labor market clearing condition (11) yields

$$
Y(t) = L(t)A(t) \left(N(t)Q(t)\right)^{\frac{\beta}{1-\beta}}.
$$

We then use (11) to replace $L(t)A(t)$ with $\left(\frac{\bar{w}(t)}{1-\beta}\right)^{-\frac{1}{\beta}} N(t)Q(t)Y(t)^{-\frac{\beta}{1-\beta}}$. Thus

$$
Y(t) = \left(\frac{\bar{w}(t)}{1-\beta}\right)^{-\frac{1-\beta}{\beta}} N(t)Q(t).
$$

On a BGP, $E(t) = \mu_e N(t)\phi Q(t)$. Therefore

$$
\frac{E(t)}{Y(t)} = \mu_e \phi \left(\frac{\bar{w}(t)}{1-\beta}\right)^{\frac{1-\beta}{\beta}}
$$

The aggregate cost of intensive and extensive margin investment is given by:

$$
\begin{aligned}
R(t) &= \sum_{\tau} \left(h_X^{\tau}(z_X^{\tau}) + h_I^{\tau}(z_I^{\tau})\right) \int_{\mathcal{N}_{\tau}(t)} q_j(t) dj \\
&= \sum_{\tau} \left(h_X^{\tau}(z_X^{\tau}) + h_I^{\tau}(z_I^{\tau})\right) N_{\tau}(t) Q_{\tau}(t) \\
&= N(t)Q(t) \sum_{\tau} \left(h_X^{\tau}(z_X^{\tau}) + h_I^{\tau}(z_I^{\tau})\right) s_{\tau},
\end{aligned}
$$

where s_{τ} is defined in (14). As a result

$$
\frac{R(t)}{Y(t)} = \left(\sum_{\tau} \left(h_X^{\tau}(z_X^{\tau}) + h_I^{\tau}(z_I^{\tau})\right) s_{\tau}\right) \left(\frac{\bar{w}(t)}{1-\beta}\right)^{\frac{1-\beta}{\beta}}.
$$

Combining the expressions for the ratios E/Y and R/Y, we obtain the fraction of total investment over output in (36).

B Distributional analyses

B.1 Derivations of Kolmogorov equations

First notice that $\bar{\mathcal{M}}_\tau(n) = \mathcal{M}_\tau(n,0)$, therefore (21) is a special case of (23) with $\hat{q}=0$. To derive the latter, let $\hat{\mathcal{M}}_{n,t,q}$ denote the measure of firms with n establishments and with each establishment having quality of at least $\hat{q}Q_t$. Using standard continuous time manipulation of Poisson processes, we have

$$\begin{aligned}\hat{\mathcal{M}}_\tau\left(1,\hat{q};t+\Delta t\right) = {} & \hat{\mathcal{M}}_\tau\left(1,\frac{\hat{q}}{\exp\left((z_I^\tau-\zeta)\Delta t\right)};t\right) - (z_X^\tau+\delta_\tau+d_\tau)\Delta t\hat{\mathcal{M}}_\tau\left(1,\frac{\hat{q}}{\exp\left((z_I^\tau-\zeta)\Delta t\right)};t\right)\\ & + 2\delta_\tau\Delta t\hat{\mathcal{M}}_\tau\left(2,\frac{\hat{q}}{\exp\left((z_I^\tau-\zeta)\Delta t\right)};t\right) + \mu_e m_\tau \Delta t N_t\left(1-\Phi\left(\frac{\hat{q}}{\exp\left((z_I^\tau-\zeta)\Delta t\right)}\right)\right)\\ & + \sum_{\tau'\neq\tau}\lambda_{\tau'\tau}\Delta t\hat{\mathcal{M}}_{\tau'}\left(1,\frac{\hat{q}}{\exp\left((z_I^\tau-\zeta)\Delta t\right)};t\right) - \sum_{\tau'\neq\tau}\lambda_{\tau\tau'}\Delta t\hat{\mathcal{M}}_\tau\left(1,\frac{\hat{q}}{\exp\left((z_I^\tau-\zeta)\Delta t\right)};t\right).\end{aligned}$$

Subtracting $\hat{\mathcal{M}}_\tau\left(n,\hat{q};t\right)$ from both sides, dividing by Δt, and take the limit $\Delta t\to 0$, we obtain

$$\begin{aligned}\frac{\partial\hat{\mathcal{M}}_\tau\left(1,\hat{q};t\right)}{\partial t} = {} & -\hat{q}(z_I^\tau-\zeta)\frac{\partial\hat{\mathcal{M}}_\tau\left(1,\hat{q};t\right)}{\partial\hat{q}} - (z_X^\tau+\delta_\tau+d_\tau)\hat{\mathcal{M}}_\tau\left(1,\hat{q};t\right)\\ & + 2\delta_\tau\hat{\mathcal{M}}_\tau\left(2,\frac{\hat{q}}{\exp\left((z_I^\tau-\zeta)\Delta t\right)};t\right) + \mu_e m_\tau N_t\left(1-\Phi\left(\hat{q}\right)\right)\\ & + \sum_{\tau'\neq\tau}\lambda_{\tau'\tau}\hat{\mathcal{M}}_{\tau'}\left(1,\hat{q};t\right) - \sum_{\tau'\neq\tau}\lambda_{\tau\tau'}\hat{\mathcal{M}}_\tau\left(1,\hat{q};t\right).\end{aligned}$$

Now $\mathcal{M}_\tau(1,\hat{q};t) = \frac{\hat{\mathcal{M}}_\tau(1,\hat{q};t+\Delta t)}{N_t}$, thus

$$\begin{aligned}\frac{\partial\mathcal{M}_\tau\left(1,\hat{q};t\right)}{\partial t} + \eta\mathcal{M}_\tau\left(1,\hat{q};t\right) = {} & -\hat{q}(z_I^\tau-\zeta)\frac{\partial\mathcal{M}_\tau\left(1,\hat{q};t\right)}{\partial\hat{q}} - (z_X^\tau+\delta_\tau+d_\tau)\mathcal{M}_\tau\left(1,\hat{q};t\right)\\ & + 2\delta_\tau\mathcal{M}_\tau\left(2,\frac{\hat{q}}{\exp\left((z_I^\tau-\zeta)\Delta t\right)};t\right) + \mu_e m_\tau\left(1-\Phi\left(\hat{q}\right)\right)\\ & + \sum_{\tau'\neq\tau}\lambda_{\tau'\tau}\mathcal{M}_{\tau'}\left(1,\hat{q};t\right) - \sum_{\tau'\neq\tau}\lambda_{\tau\tau'}\mathcal{M}_\tau\left(1,\hat{q};t\right).\end{aligned}$$

On a stationary BGP, $\frac{\partial\mathcal{M}_\tau(1,\hat{q};t)}{\partial t}=0$, so we obtain (23a).

Similarly, for $n > 1$,

$$\begin{aligned}
\hat{\mathcal{M}}_\tau\left(n, \hat{q}; t + \Delta t\right) = {} & \hat{\mathcal{M}}_\tau\left(n, \frac{\hat{q}}{\exp\left((z_I^\tau - \zeta)\Delta t\right)}; t\right) - (z_X^\tau + \delta_\tau + d_\tau)\Delta t \hat{\mathcal{M}}_\tau\left(n, \frac{\hat{q}}{\exp\left((z_I^\tau - \zeta)\Delta t\right)}; t\right) \\
& + (n+1)\delta_\tau \Delta t \hat{\mathcal{M}}_\tau\left(n+1, \frac{\hat{q}}{\exp\left((z_I^\tau - \zeta)\Delta t\right)}; t\right) \\
& + (n-1) z_X^\tau \Delta t \hat{\mathcal{M}}_\tau\left(n-1, \frac{\hat{q}}{\exp\left((z_I^\tau - \zeta)\Delta t\right)}; t\right) \\
& + \sum_{\tau' \neq \tau} \lambda_{\tau'\tau} \Delta t \hat{\mathcal{M}}_{\tau'}\left(n, \frac{\hat{q}}{\exp\left((z_I^\tau - \zeta)\Delta t\right)}; t\right) - \sum_{\tau' \neq \tau} \lambda_{\tau\tau'} \Delta t \hat{\mathcal{M}}_\tau\left(n, \frac{\hat{q}}{\exp\left((z_I^\tau - \zeta)\Delta t\right)}; t\right).
\end{aligned}$$

Subtracting $\hat{\mathcal{M}}_\tau\left(n, \hat{q}; t\right)$ from both sides, dividing by Δt, and take the limit $\Delta t \to 0$, we obtain

$$\begin{aligned}
\frac{\partial \hat{\mathcal{M}}_\tau\left(n, \hat{q}; t\right)}{\partial t} = {} & -\hat{q}(z_I^\tau - \zeta)\frac{\partial \hat{\mathcal{M}}_\tau\left(n, \hat{q}; t\right)}{\partial \hat{q}} - (z_X^\tau + \delta_\tau + d_\tau)\hat{\mathcal{M}}_\tau\left(n, \hat{q}; t\right) \\
& + (n+1)\delta_\tau \hat{\mathcal{M}}_\tau\left(n+1, \frac{\hat{q}}{\exp\left((z_I^\tau - \zeta)\Delta t\right)}; t\right) + (n-1) z_X^\tau \hat{\mathcal{M}}_\tau\left(n-1, \hat{q}; t\right) \\
& + \sum_{\tau' \neq \tau} \lambda_{\tau'\tau} \hat{\mathcal{M}}_{\tau'}\left(n, \hat{q}; t\right) - \sum_{\tau' \neq \tau} \lambda_{\tau\tau'} \hat{\mathcal{M}}_\tau\left(n, \hat{q}; t\right).
\end{aligned}$$

Since $\mathcal{M}_\tau(n, \hat{q}; t) = \frac{\hat{\mathcal{M}}_\tau(n,\hat{q};t+\Delta t)}{N_t}$, the last equality implies

$$\begin{aligned}
\frac{\partial \mathcal{M}_\tau\left(n, \hat{q}; t\right)}{\partial t} + \eta \mathcal{M}_\tau\left(n, \hat{q}; t\right) = {} & -\hat{q}(z_I^\tau - \zeta)\frac{\partial \mathcal{M}_\tau\left(n, \hat{q}; t\right)}{\partial \hat{q}} - (z_X^\tau + \delta_\tau + d_\tau)\mathcal{M}_\tau\left(1, \hat{q}; t\right) \\
& + (n+1)\delta_\tau \mathcal{M}_\tau\left(n+1, \frac{\hat{q}}{\exp\left((z_I^\tau - \zeta)\Delta t\right)}; t\right) + (n-1) z_X^\tau \mathcal{M}_\tau\left(n-1, \hat{q}; t\right) \\
& + \sum_{\tau' \neq \tau} \lambda_{\tau'\tau} \mathcal{M}_{\tau'}\left(n, \hat{q}; t\right) - \sum_{\tau' \neq \tau} \lambda_{\tau\tau'} \mathcal{M}_\tau\left(n, \hat{q}; t\right).
\end{aligned}$$

On a stationary BGP, $\frac{\partial \mathcal{M}_\tau(n,\hat{q};t)}{\partial t} = 0$, so we obtain (23b).

The derivation of (22) follows closely Cao and Luo (2017) for wealth distribution with persistent heterogeneous returns.

B.2 Proofs

Proof of Proposition 2.

First, we show that the distribution of establishment sizes has Pareto tail with the tail index given by (25). We rewrite (24) as

$$e^{\frac{\delta+d+\eta-z_X}{z_I-\zeta}p}\tilde{\mathcal{H}}(p) = \int_{-\infty}^{p} e^{\frac{\delta+d+\eta-z_X}{z_I-\zeta}\tilde{p}} \frac{\delta+d+\eta-z_X}{z_I-\zeta}(1-\Phi(\exp(\tilde{p})))d\tilde{p}.$$

Because Φ has thin tail,

$$1 - \Phi(\exp(\tilde{p})) < Ae^{-2\frac{\delta+d+\eta-z_X}{z_I-\zeta}\tilde{p}}$$

for some $A > 0$ and for all $\tilde{p} > 0$, which implies

$$a = \int_{-\infty}^{\infty} e^{\frac{\delta+d+\eta-z_X}{z_I-\zeta}\tilde{p}} \frac{\delta + d + \eta - z_X}{z_I - \zeta}(1 - \Phi(\exp(\tilde{p})))d\tilde{p} < \infty.$$

Therefore

$$\lim_{p\to\infty} e^{\frac{\delta+d+\eta-z_X}{z_I-\zeta}p}\tilde{\mathcal{H}}(p) = a.$$

That is, $\mathcal{H}$ has Pareto tail with tail index given by (25).

The proof for the distribution of the number of establishments per firm is substantially more complicated. By (18), $\frac{d+\eta}{z_X-\delta} > 1$. In this appendix, we provide the proof for the case

$$\frac{d+\eta}{z_X-\delta} < 2.$$

The proof for the case with higher $\frac{d+\eta}{z_X-\delta}$ is similar but with much more algebras.

To prove the result, we show a slightly stronger one:

$$\bar{M}(n) \propto_{n\to\infty} n^{-\frac{d+\eta}{z_X-\delta}-1}. \tag{38}$$

To do so we use the probability generating function

$$\mathcal{P}(\omega) = \sum_{n=1}^{\infty} \bar{M}(n)\omega^n.$$

The Karamata Tauberian theorem for power series Bingham et al. (1987) allows us to establish the asymptotic behavior of the cumulative sum of $\bar{M}(n)$ $(n \to \infty)$ from the asymptotic behavior of $\mathcal{P}(\omega)$ $(\omega \to 1)$ if the latter diverges. However, $\mathcal{P}(1) = 1$ so the theorem does not directly apply. In order to apply the theorem, we need to work with $\mathcal{P}''(\omega)$. Lemma 1 below provides us with the asymptotic behavior of $\mathcal{P}''(\omega)$ $(\omega \to 1)$. By the Karamata Tauberian theorem for power series (Bingham et al., 1987, Corollary 1.7.3),

$$\sum_{k=0}^{n}(k+2)(k+1)\bar{M}(k+2) \propto_{n\to\infty} n^{2-\frac{d+\eta}{z_X-\delta}}. \tag{39}$$

Now we use this result to prove (38).

Differentiating $\mathcal{P}(\omega)$ with respect to ω, we obtain

$$\mathcal{P}'(\omega) = \sum_{n=1}^{\infty} \bar{M}(n)n\omega^{n-1} = \sum_{n=0}^{\infty} \bar{M}(n+1)(n+1)\omega^n.$$

This implies

$$\omega \mathcal{P}'(\omega) = \sum_{n=1}^{\infty} \bar{M}(n) n \omega^n$$

and

$$\omega^2 \mathcal{P}'(\omega) = \sum_{n=1}^{\infty} \bar{M}(n) n \omega^{n+1} = \sum_{n=2}^{\infty} \bar{M}(n-1)(n-1)\omega^n.$$

Therefore

$$\begin{aligned}
& z_X \mathcal{P}'(\omega)\omega^2 - (z_X + \delta)\omega \mathcal{P}'(\omega) + \delta \mathcal{P}'(\omega) \\
&= \sum_{n=2}^{\infty} \left(z_X \bar{M}(n-1)(n-1) - n(z_X + \delta)\bar{M}(n)n + \delta \bar{M}(n+1)(n+1) \right) \omega^n \\
&- (z_X + \delta)\bar{M}(1)\omega + \delta \bar{M}(2)2\omega + \delta \bar{M}(1).
\end{aligned}$$

Using equalities (26b), the last equation is equivalent to

$$\begin{aligned}
& z_X \mathcal{P}'(\omega)\omega^2 - (z_X + \delta)\omega \mathcal{P}'(\omega) + \delta \mathcal{P}'(\omega) \\
&= \sum_{n=2}^{\infty} (d + \eta) \bar{M}(n)\omega^n - (z_X + \delta)\bar{M}(1)\omega + \delta \bar{M}(2)2\omega + \delta \bar{M}(1) \\
&= (d + \eta)\mathcal{P}(\omega) - \left((d + \eta + z_X + \delta)\bar{M}(1) - 2\delta \bar{M}(2) \right) \omega + \delta \bar{M}(1).
\end{aligned}$$

Rearranging and regrouping different terms and using (26a), we arrive at

$$\mathcal{P}'(\omega) \left(\delta + z_X \omega^2 - (z_X + \delta)\omega \right) = (d + \eta)\mathcal{P}(\omega) - \mu_e \omega + \delta \bar{M}(1). \tag{40}$$

Differentiating both sides twice and rearranging terms we obtain

$$\begin{aligned}
& \mathcal{P}'''(\omega) \left(\delta + z_X \omega^2 - (z_X + \delta)\omega \right) \\
&= \left(d + \eta + 2(z_X + \delta) - 4 z_X \omega \right) \mathcal{P}''(\omega) - \mathcal{P}'(\omega) 2 z_X - \mu_e.
\end{aligned} \tag{41}$$

Dividing both sides by $\delta + z_X \omega^2 - (z_X + \delta)\omega$ and observing that

$$\frac{1}{\delta + z_X \omega^2 - (z_X + \delta)\omega} = -\frac{1}{z_X - \delta} \left(\frac{1}{\omega - \frac{\delta}{z_X}} + \frac{1}{1 - \omega} \right), \tag{42}$$

we rewrite (41) as

$$
\begin{aligned}
\mathcal{P}'''(\omega) = & -\left(d+\eta+2(z_X+\delta)-4z_X\omega\right)\mathcal{P}''(\omega)\frac{1}{z_X-\delta}\left(\frac{1}{\omega-\delta/z_X}+\frac{1}{1-\omega}\right) \\
& + \mathcal{P}'(\omega)2z_X\frac{1}{z_X-\delta}\left(\frac{1}{\omega-\delta/z_X}+\frac{1}{1-\omega}\right) \\
& + \mu_e\frac{1}{z_X-\delta}\left(\frac{1}{\omega-\delta/z_X}+\frac{1}{1-\omega}\right).
\end{aligned}
$$

Equivalently,

$$
\begin{aligned}
\mathcal{P}'''(\omega) = & \left\{-\left(d+\eta+2(z_X+\delta)-4z_X\omega\right)\mathcal{P}''(\omega)+\mathcal{P}'(\omega)2z_X+\mu_e\right\}\frac{1}{z_X-\delta}\left(\frac{1}{\omega-\delta/z_X}-\frac{1}{1-\delta/z_X}\right) \\
& + \left\{-\left(d+\eta+2(z_X+\delta)-4z_X\omega\right)\mathcal{P}''(\omega)+\mathcal{P}'(\omega)2z_X+\mu_e\right\}\frac{1}{z_X} \\
& + \mathcal{P}'(\omega)2z_X\frac{1}{z_X-\delta}\frac{1}{1-\omega}+\mu_e\frac{1}{z_X-\delta}\frac{1}{1-\omega}-4z_X\mathcal{P}''(\omega)\frac{1}{z_X-\delta} \\
& + \left(2-\frac{d+\eta}{z_X-\delta}\right)\mathcal{P}''(\omega)\frac{1}{1-\omega}. \qquad (43)
\end{aligned}
$$

Let $\mathcal{Q}$ be defined by

$$
\mathcal{Q}(\omega) \equiv \left\{-\left(d+\eta+2(z_X+\delta)-4z_X\omega\right)\mathcal{P}''(\omega)+\mathcal{P}'(\omega)2z_X+\mu_e\right\}\frac{1}{z_X-\delta}\left(\frac{1}{\omega-\delta/z_X}-\frac{1}{1-\delta/z_X}\right).
$$

It follows that $\mathcal{Q}(\omega)$ is finite for all $\omega < 1$. Lemma 1 and the fact that

$$
\frac{1}{\omega-\delta/z_X}-\frac{1}{1-\delta/z_X} = O(1-\omega)
$$

imply

$$
\lim_{\omega\to 1}\mathcal{Q}(\omega) = 0,
$$

when $\omega \to 1$. Therefore, by the Riemann-Lebsegue lemma, the Taylor expansion of $\mathcal{Q}(\omega)$

$$
\mathcal{Q}(\omega) = \sum_{n=0}^{\infty} q_n\omega^n
$$

satisfies

$$
\lim_{n\to\infty} q_n = 0.
$$

By comparing the power series for both sides of (43), coefficient by coefficient, we obtain

$$
\begin{aligned}
&(n+3)(n+2)(n+1)\bar{M}(n+3) \\
&= q_n + (d+\eta+2(z_X+\delta))\,(n+2)(n+1)\bar{M}(n+2) - 4z_X(n+1)\bar{M}(n+1) + 2z_X(n+1)\bar{M}(n+1) \\
&+ \frac{2z_X}{z_X-\delta}\sum_{k=0}^{n}(k+1)\bar{M}(k+1) + \frac{\mu_e}{z_X-\delta} - \frac{4z_X}{z_X-\delta}(n+2)(n+1)\bar{M}(n) \\
&+ \left(2 - \frac{d+\eta}{z_X-\delta}\right)\sum_{k=0}^{n}(k+2)(k+1)\bar{M}(k+2)
\end{aligned}
$$

Observing that $(n+3)(n+2)(n+1)\bar{M}(n)$ is the leading term on the right hand side of the last expression, so (39) implies

$$(n+3)(n+2)(n+1)\bar{M}(n+3) \propto_{n\to\infty} n^{2-\frac{d+\eta}{z_X-\delta}}.$$

This limit is equivalent to (38).

Now (38) together with Lemma 2 yields

$$\Pr(\mathbf{X} \geq n) = \sum_{k\geq n}\bar{M}(k) \propto n^{-\frac{d+\eta}{z_X-\delta}},$$

as $n \to \infty$. ■

Lemma 1 *Assume* $\frac{d+\eta}{z_X-\delta} \in (1,2)$, *then the second derivative of the probability generating function satisfies*

$$\mathcal{P}''(\omega) \propto (1-\omega)^{\frac{d+\eta}{z_X-\delta}-2}$$

as $\omega \to 1$.

Proof. Dividing both sides of (40) and using identity (42), we arrive at

$$\mathcal{P}'(\omega) + \mathcal{P}(\omega)\frac{d+\eta}{z_X-\delta}\left(\frac{1}{\omega-\frac{\delta}{z_X}} + \frac{1}{1-\omega}\right) = \frac{\mu_e\omega - \delta\bar{M}(1)}{z_X-\delta}\left(\frac{1}{\omega-\frac{\delta}{z_X}} + \frac{1}{1-\omega}\right).$$

Let

$$\psi(\omega) = \left(\frac{\omega-\frac{\delta}{z_X}}{1-\omega}\right)^{\frac{d+\eta}{z_X-\delta}}$$

which satisfies

$$\psi'(\omega) = \frac{d+\eta}{z_X-\delta}\left(\frac{1}{\omega-\frac{\delta}{z_X}} + \frac{1}{1-\omega}\right)\psi(\omega). \tag{44}$$

Then the differential equation for $\mathcal{P}(\omega)$ above can be rewritten as

$$\frac{d}{d\omega}\left(\mathcal{P}(\omega)\psi(\omega)\right) = \mathcal{P}'(\omega)\psi(\omega) + \mathcal{P}(\omega)\psi'(\omega) = \psi(\omega)\frac{\mu_e\omega - \delta\bar{M}(1)}{z_X-\delta}\left(\frac{1}{\omega-\frac{\delta}{z_X}} + \frac{1}{1-\omega}\right).$$

Integrating both sides from some $\underline{\omega} > \delta/z_X$ up to any $\omega \in (\underline{\omega}, 1)$:

$$\begin{aligned}
\mathcal{P}(\omega)\psi(\omega) &= \mathcal{P}(\underline{\omega})\psi(\underline{\omega}) + \int_{\underline{\omega}}^{\omega} \psi(\tilde{\omega}) \frac{\mu_e\tilde{\omega} - \delta\bar{M}(1)}{z_X - \delta}\left(\frac{1}{\tilde{\omega} - \frac{\delta}{z_X}} + \frac{1}{1-\tilde{\omega}}\right) d\tilde{\omega} \\
&= \mathcal{P}(\underline{\omega})\psi(\underline{\omega}) + \int_{\underline{\omega}}^{\omega} \psi'(\tilde{\omega}) \frac{\mu_e\tilde{\omega} - \delta\bar{M}(1)}{d+\eta} d\tilde{\omega},
\end{aligned}$$

where the second equality is due to (44). Equivalently,

$$\mathcal{P}(\omega) = \frac{c}{\psi(\omega)} + \frac{1}{\psi(\omega)} \int_{\underline{\omega}}^{\omega} \psi'(\tilde{\omega}) \frac{\mu_e\tilde{\omega} - \delta\bar{M}(1)}{d+\eta} d\tilde{\omega},$$

where $c = \mathcal{P}(\underline{\omega})\psi(\underline{\omega}) > 0$.

Integrating by parts, we obtain

$$\int_{\underline{\omega}}^{\omega} \psi'(\tilde{\omega}) \frac{\mu_e\tilde{\omega} - \delta\bar{M}(1)}{d+\eta} d\tilde{\omega} = \psi(\omega)\frac{\mu_e\omega - \delta\bar{M}(1)}{d+\eta} - \psi(\underline{\omega})\frac{\mu_e\underline{\omega} - \delta\bar{M}(1)}{d+\eta} - \int_{\underline{\omega}}^{\omega} \psi(\tilde{\omega})\frac{\mu_e}{d+\eta} d\tilde{\omega}.$$

Therefore

$$\mathcal{P}(\omega) = \frac{c - \psi(\underline{\omega})\frac{\mu_e\underline{\omega} - \delta\bar{M}(1)}{d+\eta}}{\psi(\omega)} + \frac{\mu_e\omega - \delta\bar{M}(1)}{d+\eta} - \frac{\int_{\underline{\omega}}^{\omega} \psi(\tilde{\omega})\frac{\mu_e}{d+\eta} d\tilde{\omega}}{\psi(\omega)}.$$

The derivatives can be computed explicitly:

$$\begin{aligned}
\mathcal{P}'(\omega) &= -\frac{c - \psi(\underline{\omega})\frac{\mu_e\underline{\omega} - \delta\bar{M}(1)}{d+\eta}}{\psi(\omega)}\frac{\psi'(\omega)}{\psi(\omega)} + \frac{\int_{\underline{\omega}}^{\omega} \psi(\tilde{\omega})\frac{\mu_e}{d+\eta} d\tilde{\omega}}{\psi(\omega)}\frac{\psi'(\omega)}{\psi(\omega)} \\
&= -\frac{c - \psi(\underline{\omega})\frac{\mu_e\underline{\omega} - \delta\bar{M}(1)}{d+\eta}}{\psi(\omega)}\frac{d+\eta}{z_X - \delta}\left(\frac{1}{\omega - \frac{\delta}{z_X}} + \frac{1}{1-\omega}\right) \\
&\quad + \frac{\int_{\underline{\omega}}^{\omega} \psi(\tilde{\omega})\frac{\mu_e}{d+\eta} d\tilde{\omega}}{\psi(\omega)}\frac{d+\eta}{z_X - \delta}\left(\frac{1}{\omega - \frac{\delta}{z_X}} + \frac{1}{1-\omega}\right)
\end{aligned}$$

and

$$\begin{aligned}
&\mathcal{P}''(\omega) \\
&= \frac{c - \psi(\underline{\omega})\frac{\mu_e\underline{\omega} - \delta\bar{M}(1)}{d+\eta}}{\psi(\omega)}\left\{\left(\frac{d+\eta}{z_X - \delta}\right)^2\left(\frac{1}{\omega - \frac{\delta}{z_X}} + \frac{1}{1-\omega}\right)^2 - \frac{d+\eta}{z_X - \delta}\left(-\frac{1}{(\omega - \frac{\delta}{z_X})^2} + \frac{1}{(1-\omega)^2}\right)\right\} \\
&+ \mathcal{R}(\omega)
\end{aligned}$$

where

$$
\begin{aligned}
\mathcal{R}(\omega) = {} & \frac{d+\eta}{z_X-\delta}\left(-\frac{1}{(\omega-\frac{\delta}{z_X})^2}+\frac{1}{(1-\omega)^2}\right)\frac{\int_{\underline{\omega}}^{\omega}\psi(\tilde{\omega})\frac{\mu_e}{d+\eta}d\tilde{\omega}}{\psi(\omega)} \\
& -\left(\frac{d+\eta}{z_X-\delta}\right)^2\left(\frac{1}{(\omega-\frac{\delta}{z_X})^2}+\frac{2}{(\omega-\frac{\delta}{z_X})(1-\omega)}+\frac{1}{(1-\omega)^2}\right)\frac{\int_{\underline{\omega}}^{\omega}\psi(\tilde{\omega})\frac{\mu_e}{d+\eta}d\tilde{\omega}}{\psi(\omega)} \\
& +\frac{d+\eta}{z_X-\delta}\left(\frac{1}{\omega-\frac{\delta}{z_X}}+\frac{1}{1-\omega}\right)\frac{\mu_e}{d+\eta}.
\end{aligned}
$$

Now,

$$
\begin{aligned}
& \int_{\underline{\omega}}^{\omega}\psi(\tilde{\omega})d\tilde{\omega} = \int_{\underline{\omega}}^{\omega}\left(\tilde{\omega}-\frac{\delta}{z_X}\right)^{\frac{d+\eta}{z_X-\delta}}(1-\tilde{\omega})^{-\frac{d+\eta}{z_X-\delta}}d\tilde{\omega} \\
& = \int_{\underline{\omega}}^{\omega}\left(\tilde{\omega}-\frac{\delta}{z_X}\right)^{\frac{d+\eta}{z_X-\delta}} d\left(\frac{(1-\tilde{\omega})^{-\frac{d+\eta}{z_X-\delta}+1}}{\frac{d+\eta}{z_X-\delta}-1}\right) \\
& = \left.\frac{\left(\tilde{\omega}-\frac{\delta}{z_X}\right)^{\frac{d+\eta}{z_X-\delta}}(1-\tilde{\omega})^{-\frac{d+\eta}{z_X-\delta}+1}}{\frac{d+\eta}{z_X-\delta}-1}\right|_{\underline{\omega}}^{\omega} - \int_{\underline{\omega}}^{\omega}\frac{d+\eta}{z_X-\delta}\left(\tilde{\omega}-\frac{\delta}{z_X}\right)^{\frac{d+\eta}{z_X-\delta}-1}\frac{1}{\frac{d+\eta}{z_X-\delta}-1}(1-\tilde{\omega})^{-\frac{d+\eta}{z_X-\delta}+1}d\tilde{\omega} \\
& = \frac{\left(\omega-\frac{\delta}{z_X}\right)^{\frac{d+\eta}{z_X-\delta}}(1-\omega)^{-\frac{d+\eta}{z_X-\delta}+1}}{\frac{d+\eta}{z_X-\delta}-1}+c_\psi+O(1-\omega)
\end{aligned}
$$

where

$$
c_\psi = -\frac{\psi(\underline{\omega})(1-\underline{\omega})}{\frac{d+\eta}{z_X-\delta}-1}-\int_{\underline{\omega}}^{1}\frac{d+\eta}{z_X-\delta}\left(\tilde{\omega}-\frac{\delta}{z_X}\right)^{\frac{d+\eta}{z_X-\delta}-1}\frac{1}{\frac{d+\eta}{z_X-\delta}-1}(1-\tilde{\omega})^{-\frac{d+\eta}{z_X-\delta}+1}d\tilde{\omega}.
$$

So

$$
\frac{\int_{\underline{\omega}}^{\omega}\psi(\tilde{\omega})d\tilde{\omega}}{\psi(\omega)} = \frac{1-\omega}{\frac{d+\eta}{z_X-\delta}-1}+\frac{c_\psi}{\psi(\omega)}+o(1-\omega)
$$

Therefore the factor associated with $\frac{1}{1-\omega}$ in $\mathcal{R}(\omega)$ is

$$
\left(\frac{d+\eta}{z_X-\delta}-\left(\frac{d+\eta}{z_X-\delta}\right)^2\right)\frac{1}{\frac{d+\eta}{z_X-\delta}-1}\frac{\mu_e}{d+\eta}+\frac{d+\eta}{z_X-\delta}\frac{\mu_e}{d+\eta}=0.
$$

which implies

$$
\mathcal{R}(\omega) = O(1)+\frac{c_\psi}{\psi(\omega)(1-\omega)^2}\frac{d+\eta}{z_X-\delta}\left(1-\frac{d+\eta}{z_X-\delta}\right)
$$

as $\omega \to 1$. So

$$\mathcal{P}''(\omega) \sim_{\omega\to 1} \left(c - \psi(\underline{\omega})\frac{\mu_e\underline{\omega} - \delta\bar{M}(1)}{d+\eta} - c_\psi\right)\frac{d+\eta}{z_X-\delta}\left(\frac{d+\eta}{z_X-\delta} - 1\right)\frac{1}{\psi(\omega)(1-\omega)^2}$$

Notice that

$$c - \psi(\underline{\omega})\frac{\mu_e\underline{\omega} - \delta\bar{M}(1)}{d+\eta} - c_\psi > \psi(\underline{\omega})\left(\mathcal{P}(\underline{\omega}) - \frac{\mu_e\underline{\omega} - \delta\bar{M}(1)}{d+\eta} + \frac{1-\underline{\omega}}{\frac{d+\eta}{z_X-\delta} - 1}\right)$$

and

$$\begin{aligned}
&\mathcal{P}(\underline{\omega}) - \frac{\mu_e\underline{\omega} - \delta\bar{M}(1)}{d+\eta} + \frac{1-\underline{\omega}}{\frac{d+\eta}{z_X-\delta} - 1} \\
&= \frac{\mathcal{P}'(\underline{\omega})\left(\delta + z_X\underline{\omega}^2 - (z_X+\delta)\underline{\omega}\right)}{d+\eta} + \frac{1-\underline{\omega}}{\frac{d+\eta}{z_X-\delta} - 1} \\
&= \left\{\frac{\mathcal{P}'(\underline{\omega})(\delta - z_X\underline{\omega})}{d+\eta} + \frac{1}{\frac{d+\eta}{z_X-\delta} - 1}\right\}(1-\underline{\omega}) > 0,
\end{aligned}$$

when $\underline{\omega}$ is chosen sufficiently close to δ/z_X. Thus

$$c - \psi(\underline{\omega})\frac{\mu_e\underline{\omega} - \delta\bar{M}(1)}{d+\eta} - c_\psi > 0.$$

■

Lemma 2 *Suppose that*

$$\bar{M}(n) \propto_{n\to\infty} n^{-\frac{d+\eta}{z_X-\delta}-1},$$

then

$$\Pr(\mathbf{X} \geq n) = \sum_{k\geq n} \bar{M}(k) \propto n^{-\frac{d+\eta}{z_X-\delta}}.$$

Proof. There exists $a > 0$ such that

$$\lim_{n\to\infty} \frac{\bar{M}(n)}{n^{-\frac{d+\eta}{z_X-\delta}-1}.} = a.$$

Therefore, for any $\epsilon > 0$, there exists n^* such that, for all $n \geq n^*$

$$a - \epsilon < \frac{\bar{M}(n)}{n^{-\frac{d+\eta}{z_X-\delta}-1}.} < a + \epsilon$$

Combining these inequalities with the definition of $\Pr(\mathbf{X} \geq n)$, we obtain, for all $n \geq n^*$

$$(a-\epsilon)\sum_{k\geq n} k^{-\frac{d+\eta}{z_X-\delta}-1} < \Pr(\mathbf{X} \geq n) = \sum_{k\geq n} \bar{M}(k) < (a+\epsilon)\sum_{k\geq n} k^{-\frac{d+\eta}{z_X-\delta}-1}.$$

Notice that

$$\sum_{k\geq n} k^{-\frac{d+\eta}{z_X-\delta}-1} < \int_{k\geq n-1} k^{-\frac{d+\eta}{z_X-\delta}-1} dk = \frac{z_X-\delta}{d+\eta}(n-1)^{-\frac{d+\eta}{z_X-\delta}}$$

and

$$\sum_{k\geq n} k^{-\frac{d+\eta}{z_X-\delta}-1} > \int_{k\geq n} k^{-\frac{d+\eta}{z_X-\delta}-1} dk = \frac{z_X-\delta}{d+\eta} n^{-\frac{d+\eta}{z_X-\delta}}.$$

Therefore

$$(a-\epsilon)\frac{z_X-\delta}{d+\eta} n^{-\frac{d+\eta}{z_X-\delta}} < \Pr(\mathbf{X}\geq n) < (a+\epsilon)\frac{z_X-\delta}{d+\eta}(n-1)^{-\frac{d+\eta}{z_X-\delta}}$$

As this applies for any $\epsilon > 0$, we obtain

$$\lim_{n\to\infty} \frac{\Pr(\mathbf{X}\geq n)}{\frac{z_X-\delta}{d+\eta} n^{-\frac{d+\eta}{z_X-\delta}}} = a.$$

■

Proofs for Proposition 3. First we derive the expression for $\varphi(s)$ provide in the main text.

$$\begin{aligned}
\varphi(s) &= \int_0^\infty \hat{\mathbf{z}}^s \frac{-d\sum_n \bar{\mathcal{M}}(n)(1-\Phi(\hat{\mathbf{z}}/n))}{\sum_n \bar{\mathcal{M}}(n)} \\
&= \int_0^\infty \hat{\mathbf{z}}^s \frac{\sum_n \bar{\mathcal{M}}(n) d\Phi(\hat{\mathbf{z}}/n)}{\sum_n \bar{\mathcal{M}}(n)} \\
&= \int_0^\infty \frac{\sum_n \bar{\mathcal{M}}(n) n^s (\hat{\mathbf{z}}/n)^s d\Phi(\hat{\mathbf{z}}/n)}{\sum_n \bar{\mathcal{M}}(n)} \\
&= \left\{\int_0^\infty \hat{\mathbf{z}}^s d\Phi(\hat{\mathbf{z}})\right\}\left\{\frac{\sum_n \bar{\mathcal{M}}(n) n^s}{\sum_n \bar{\mathcal{M}}(n)}\right\}.
\end{aligned}$$

Now we prove that

$$\varphi(s) \propto \frac{1}{\frac{\eta+d}{z_X-\delta} - s} \tag{45}$$

as $s \uparrow \frac{\eta+d}{z_X-\delta}$. To do so, we use (38) which characterizes the asymptotic behavior of $\bar{M}(n)$. This result implies that, there exists $a > 0$ such that: for any $\epsilon > 0$, there exists n^* so that

$$(a-\epsilon) n^{-\frac{d+\eta}{z_X-\delta}-1} < \bar{M}(n) < (a+\epsilon) n^{-\frac{d+\eta}{z_X-\delta}-1}$$

for all $n \geq n^*$. Therefore

$$\left(\frac{\eta+d}{z_X-\delta} - s\right)\sum_{n\geq n^*}(a-\epsilon) n^{s-\frac{d+\eta}{z_X-\delta}-1} < \left(\frac{\eta+d}{z_X-\delta} - s\right)\sum_{n\geq n^*}\bar{M}(n) n^s < \left(\frac{\eta+d}{z_X-\delta} - s\right)\sum_{n\geq n^*}(a+\epsilon) n^{s-\frac{d+\eta}{z_X-\delta}-1}$$

Notice that

$$\left(\frac{\eta+d}{z_X-\delta} - s\right)\sum_{n\geq n^*} n^{s-\frac{d+\eta}{z_X-\delta}-1} < \left(\frac{\eta+d}{z_X-\delta} - s\right)\int_{n^*-1}^\infty x^{s-\frac{d+\eta}{z_X-\delta}-1} dx = (n^*-1)^{s-\frac{d+\eta}{z_X-\delta}}$$

and

$$\left(\frac{\eta+d}{z_X-\delta}-s\right)\sum_{n\geq n^*} n^{s-\frac{d+\eta}{z_X-\delta}-1} > \left(\frac{\eta+d}{z_X-\delta}-s\right)\int_{n^*}^{\infty} x^{s-\frac{d+\eta}{z_X-\delta}-1}dx = (n^*)^{s-\frac{d+\eta}{z_X-\delta}}$$

Since

$$\left(\frac{\eta+d}{z_X-\delta}-s\right)\sum_{n}\bar{M}(n)n^s = \left(\frac{\eta+d}{z_X-\delta}-s\right)\sum_{n<n^*}\bar{M}(n)n^s + \left(\frac{\eta+d}{z_X-\delta}-s\right)\sum_{n\geq n^*}\bar{M}(n)n^s,$$

using the inequalities above and take the limit $s\uparrow\frac{\eta+d}{z_X-\delta}$, we obtain

$$a-2\epsilon < \left(\frac{\eta+d}{z_X-\delta}-s\right)\sum_{n}\bar{M}(n)n^s < a+2\epsilon$$

for all $s\in(s^*,\frac{\eta+d}{z_X-\delta})$ with s^* sufficiently close to $\frac{\eta+d}{z_X-\delta}$. In other words,

$$\sum_{n}\bar{M}(n)n^s \propto_{s\uparrow\frac{\eta+d}{z_X-\delta}} \frac{1}{\frac{\eta+d}{z_X-\delta}-s}.$$

Because Φ has thin tail $\int_0^\infty \hat{\mathbf{z}}^s d\Phi(\hat{\mathbf{z}})$ is finite and continuous in s which implies (45). ■

Proof of Proposition 4. Let $\hat{\varphi}(n,s)$ denote the Laplace transform of $\mathcal{M}(n,q)$. In Subsection 5.2, we derived the difference equations satisfied by $\hat{\varphi}(n,s)$ similar to the difference equations for $\bar{M}(n)$. Using these difference equations and following the steps in the proof of Proposition 2, we can show that, there exists $a(s)$ such that

$$\lim_{n\to\infty} n^{\frac{d+\eta-(z_I-\zeta)s}{z_X-\delta}+1}\hat{\varphi}(n,s) = a(s)$$

and the convergence is uniform in s. Recall that the Laplace transformation for firm size distribution can be written as

$$\varphi(s) = \frac{\sum_n n^s\hat{\varphi}(n,s)}{\sum_n \bar{\mathcal{M}}(n)},$$

Armed with this result, we can follow the steps in the proof of Proposition 3 to show that

$$\sum_n \hat{\varphi}(n,s)n^s \propto \frac{1}{s^*-s}$$

Using this property and applying Mimica (2016, Corollary 1.3), we obtain the tail result stated in the proposition.

To derive (32), notice that, by (29),

$$d+\eta = (z_X-\delta)\lambda^{ne}$$

and by (25)

$$z_I-\zeta = \frac{d+\eta-(z_X-\delta)}{\lambda^e} = \frac{(z_X-\delta)(\lambda^{ne}-1)}{\lambda^e}.$$

Plugging these expression in (31), we obtain

$$\lambda^f = \frac{(z_X - \delta)\lambda^{ne}}{z_X - \delta + \frac{(z_X - \delta)(\lambda^{ne} - 1)}{\lambda^e}} = \frac{\lambda^{ne}\lambda^e}{\lambda^{ne} + \lambda^e - 1}.$$

Inverting the first and the last items, we arrive at (32).

■

C Robustness exercises

C.1 One-type model

To check whether two firm types are necessary, we re-estimate the model with one firm type. To do so, we first fix $\delta = 0.118$ and $d = 0.004$.[40] We target $\eta = 0.01$ and $\zeta = 0.021$ and the Pareto tail index of the distribution of number of establishments per firm of 1.25 and the Pareto tail index of the distribution of establishment size of 1.40, as estimated for 1995.

The Pareto tail index of the distribution of number of establishments per firm is given by (29). Therefore

$$z_X = \frac{\eta + d}{1.25} + \delta = 0.1294$$

From (13), we have

$$\mu_e = \eta + d + \delta - z_X = 0.0028$$

The Pareto tail index of the distribution of establishment sizes is given by (25), which implies

$$z_I = \frac{\eta + d + \delta - z_X}{1.40} + \zeta = 0.0230$$

From (17), we obtain

$$\int \hat{q} d\Phi(\hat{q}) = 1 - \frac{z_I - \zeta}{\mu_e} = 0.2857$$

We assume that Φ follows a log-normal distribution with mean ϱ and variance ς^2. Thus

$$\exp\left(\varrho + \frac{\varsigma^2}{2}\right) = \int \hat{q} d\Phi(\hat{q}) = 0.2857$$

[40] These amount to around 3% quarterly exit rate for establishments and 0.1% quarterly (exogenous) exit rate for firm.

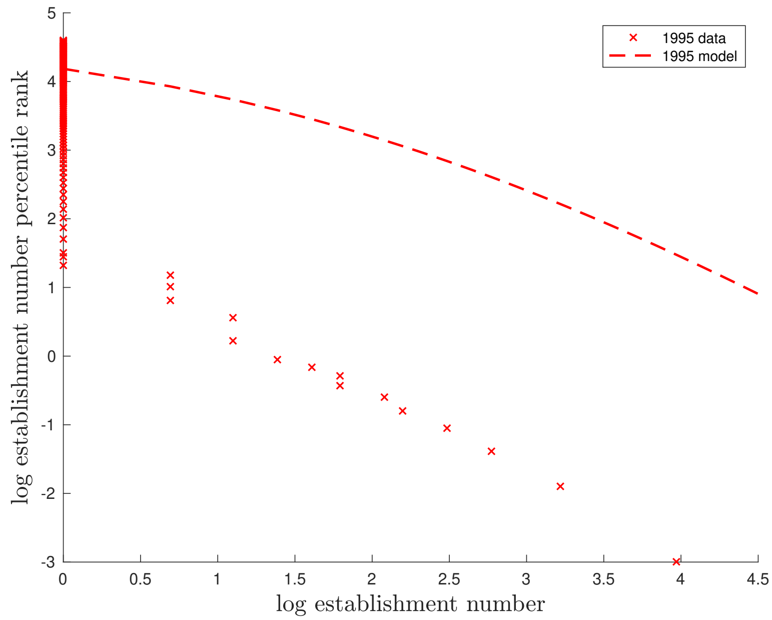


Figure C.1: Distribution of number of establishments per firm, Data and Model (One-Type)

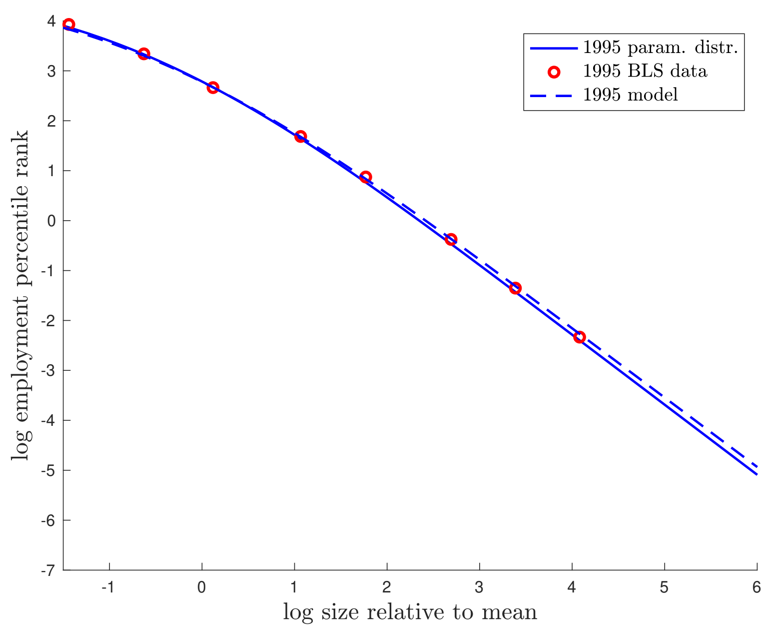


Figure C.2: Distribution of number of employees per establishment, Data and Model (One-Type)

Now, the distribution of the number of establishments per firm is given by (27). Figure C.1 displays this distribution and shows that this model produces too few firms with one establishments, despite replicating the tail index of the empirical distribution. To better match the data, we therefore need more than one type.

Notably, however, the model generates a distribution of establishment size close to the estimated parameterized distribution at the estimated values $\hat{\varsigma} = 1.1936$ and $\hat{\varrho} = -1.9652$. The model implied distribution as given by (24) and the empirical distributions are shown in Figure C.2.

C.2 Endogenous growth

In the benchmark model, our assumption implies the economy exhibits semi-endogenous growth. In this appendix, we consider alternative specifications with fully endogenous growth, so that changes in model parameters, such as the entry cost ϕ, can affect not only the level but also the growth rate of the economy. For simplicity of explication, we only provide details for parts of the model, characterization and computation that are different from the main text.

Intermediate firms: Recall that the production function for intermediate good j is $x_j(t) = A(t)\ell_j(t)$, where $A(t)$ is the labor productivity. To model semi-endogenous aggregate growth, we allow productivity to be a geometric weighted average of exogenous process $\exp(\theta t)$, and an endogenous factor, $Q(t)$, such that

$$A(t) = (e^{\theta t})^{\alpha} Q(t)^{1-\alpha}, \tag{46}$$

where

$$Q(t) \equiv \frac{1}{N(t)} \int_{\mathcal{N}(t)} q_j(t) dj$$

is the average quality of intermediate goods.

BGP : Along the BGP, the quality-invariant component of profit $\bar{\pi}(t)$ in (6) is constant. Therefore, $w(t)$ must grow at the same rate as $A(t)Y(t)^{\frac{\beta}{1-\beta}}$. Given that $Y(t)$ grows at rate g and $A(t)$ grows at rate $\alpha\theta + (1-\alpha)\zeta$ from (46), then $w(t)$ must grow at the rate $\beta g/(1-\beta) + \alpha\theta + (1-\alpha)\zeta$. This implies that the labor income of the representative consumer, $w(t)L(t)$, grows at the rate of wages plus population growth. Since labor income must grow at the same rate as consumption, we know that

$$g = \gamma + \alpha\theta + (1-\alpha)\zeta + \frac{\beta}{1-\beta} g$$

must hold.

Next, we can use the labor market clearing condition to further refine the expression for the growth rate for final output. Total labor demand can be calculated from (5),

$$\int_{\mathcal{N}(t)} \ell_j(t) dj = \left(\frac{w(t)}{(1-\beta)A(t)} \right)^{-\frac{1}{\beta}} \frac{Y(t)N(t)Q(t)}{A(t)}.$$

The labor market clearing condition $L(t) = \int_{\mathcal{N}_t} \ell_j(t)dj$ then yields the following expression for $\bar{w}(t)$,

$$\bar{w}(t) = (1-\beta)\left(\frac{N(t)Q(t)}{A(t)Y(t)^{\frac{\beta}{1-\beta}}}\right)^{\beta} L(t)^{-\beta}.$$

Since we showed that the normalized wage (or equivalently profit per quality) does not grow along the BGP, the above expression implies

$$0 = -\gamma + (\eta + \zeta) - (\alpha\theta + (1-\alpha)\zeta) - \frac{\beta}{1-\beta}g,$$

which can be combined with condition (10) to yield

$$g = \eta + \zeta.$$

Hence, the growth of aggregate output is driven by the two margins of firm growth: the growth of the number of establishments $N(t)$ and the growth of the average quality of products $Q(t)$. Furthermore, we can rewrite (10) to obtain an explicit formula for ζ given g:

$$\zeta = \frac{1}{1-\alpha}\left(\frac{1-2\beta}{1-\beta}g - \gamma - \alpha\theta\right).$$

Because $\eta = g - \zeta$,

$$\eta = g - \frac{1}{1-\alpha}\left(\frac{1-2\beta}{1-\beta}g - \gamma - \alpha\theta\right)$$

holds.

Note that using the last expression for η, (13) can be rewritten as

$$g - \frac{1}{1-\alpha}\left(\frac{1-2\beta}{1-\beta}g - \gamma - \alpha\theta\right) = z_X^{\tau} - (\delta_\tau + d_\tau) + \mu_e \frac{m_\tau}{M_\tau} - \sum_{\tau' \neq \tau} \lambda_{\tau\tau'} + \sum_{\tau' \neq \tau} \lambda_{\tau'\tau} \frac{M_{\tau'}}{M_\tau}. \tag{47}$$

One firm type characterization: The single-type case is particularly convenient as it has the recursive structure in solution. The free-entry condition (9) and (8) imply that

$$v = v_e = \phi.$$

From the first-order condition for innovations, this imply that z_I and z_X are determined only by ϕ and the innovation cost functions $h_I(\cdot)$ and $h_X(\cdot)$. Then the only endogenous variable to determine ζ and g in (17) and (19) is μ_e. Plugging (17) and (19) into (10) yields an equation with one unknown μ_e. With some algebra, this equation becomes

$$\left[\left(\frac{1-2\beta}{1-\beta} - (1-\alpha)\right)\int \hat{q}d\Phi(\hat{q}) + 1 - \alpha\right]\mu_e + \left[\frac{1-2\beta}{1-\beta} - (1-\alpha)\right]z_I + \frac{1-2\beta}{1-\beta}(z_X - \delta - d) = \gamma + \alpha\theta.$$

We assume that β and α are sufficiently small so that the coefficients of μ_e, z_I, and $(z_X - \delta - d)$ on the left-hand side are all positive. Then the above steps solve out μ_e, ζ, g, η, z_I, z_X, and v as functions of parameters. The only equilibrium object left here is $\bar{w}$, which can be solved by the HJB equation

$$(\rho + \sigma g)v = \bar{\pi} - h_I(z_I) - h_X(z_X) + (z_I + z_X - \delta - d)v$$

and the relationship (6).

Proposition 6 *For BGP equilibra in the one-type economy:*

(i) An increase in the entry cost, ϕ, increases z_I and z_X, while reducing μ_e. When $\alpha = 1$ (exogenous growth), these effects exactly offset and g stays constant. When $\alpha < 1$ and $\int \hat{q} d\Phi(\hat{q}) \leq 1$, g increases.

(ii) Suppose that the innovation cost functions take the form

$$h_i(z_i) = \chi_i z_i^{\psi},$$

where $i = I, X$. The parameters satisfy $\chi_i > 0$ and $\psi > 1$. Then a decrease in χ_I increases z_I but keeps z_X the same. The entry rate μ_e decreases, and when $\alpha < 1$, the overall growth rate increases. A decrease in χ_X increases z_X but keeps z_I the same. The entry rate μ_e decreases, and when $\alpha < 1$ and $\int \hat{q} d\Phi(\hat{q}) \leq 1$, g increases.

(iii) An increase in δ, d, γ, or θ keeps z_I and z_X constant, while increasing g through an increase in μ_e. Changes in α and β do not affect z_I and z_X either, although they influence g through the change in μ_e.

(iv) Changes in the preference parameters σ and ρ do not have any effects on z_I, z_X, and g.

Computing equilibrium: We can compute the equilibrium when $\alpha < 1$ as follows. First, make a guess on g. Once g is known, we can compute r by $r = \rho + \sigma g$ and look for $\bar{w}$ that satisfies the free-entry condition (9) by solving (7) for given r and $\bar{w}$. Then, we compare the initial guess with the g calculated from the general equilibrium. In particular, the set of equations (15), (16), (21), and (47) can be used for pinning down μ_e, g, M_τ, and s_τ ($2T + 2$ equations and $2T + 2$ unknowns). If this g is different from the initial guess, adjust g until we find the fixed point.

D Data and Empirical Documentation

D.1 BLS-QCEW Data

This data appendix describes the Quarterly Census of Employment and Wages (QCEW) and draws heavily from the BLS Handbook of Methods.[41]

D.1.1 Definitions

The Quarterly Census of Employment and Wages (QCEW) is a count of employment and wages obtained from quarterly reports filed by almost every employer in the U.S., Puerto Rico and the U.S. Virgin Islands, for the purpose of administering state unemployment insurance programs. These reports are compiled by the Bureau of Labor Statistics (BLS) and supplemented with the Annual Refiling Survey and the Multiple Worksite Report for the purpose of validation and accuracy. The reports include an establishment's monthly employment level upon the twelfth of each month and counts any employed worker, whether their position is full time, part time, permanent or temporary. Counted employees include most corporate officials, all executives, all supervisory personnel, all professionals, all clerical workers, many farmworkers, all wage earners and all piece workers. Employees are counted if on paid sick leave, paid holiday or paid vacation. Employees are not counted if they did not earn wages during the pay period covering the 12th of the month, because of work stoppages, temporary layoffs, illness, or unpaid vacations. The QCEW does not count proprietors, the unincorporated self-employed, unpaid family members, certain farm and domestic workers that are exempt from reporting employment data, railroad workers covered by the railroad unemployment insurance system, all members of the Armed Forces, and most student workers at schools. If a worker holds multiple jobs across multiple firms, then that worker may be counted more than once in the QCEW.

D.1.2 BLS Sample

A sample we used as part of the BLS visiting researcher program provided data from 1990 to 2016 and covers thirty-eight states: Alaska, Alabama, Arkansas, Arizona, California, Colorado, Connecticut, Delaware, Georgia, Hawaii, Iowa, Idaho, Indiana, Kansas, Louisiana, Maryland, Maine, Minnesota, Montana, North Dakota, New Jersey, New Mexico, Nevada, Ohio, Oklahoma, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Virginia, Vermont, Washington, West Virginia, as well as the District of Columbia, Puerto Rico and the U.S. Virgin Islands.

D.1.3 LEHD Sample

The Employer Characteristics File maintained by the Longitudinal Employer-Household Dynamics program provided data for twenty-eight states: Alaska, Arizona, California, Colorado, Florida, Georgia, Iowa, Idaho, Illinois, Indiana, Kansas, Louisiana, Maryland, Missouri, Montana, North

[41] See `https://www.bls.gov/opub/hom/cew/home.htm` for the complete BLS Handbook of Methods.

Carolina, New Mexico, New York, Pennsylvania, Oregon, Rhode Island, South Dakota, Texas, Utah, Washington, Wisconsin, West Virginia, and Wyoming.

D.1.4 Data cleaning and variable construction

To conform to official statistics, we clean the data in accordance with BLS procedure. First, while the QCEW contains monthly data as of the 12th of each month, we follow BLS convention by only using data from the final month within a quarter. As a result, our sample does not capture establishments that enter and exit within the same quarter. We additionally exclude firms from calculations in a given quarter if the absolute change in employment from the previous quarter exceeds 10 times the average employment between the two quarters. Statistics within this paper are not sensitive to the choice of multiple being 10.

We construct firms as collections of establishment with the same employment identification numbers (EINs). Firm-level employment is the sum of all employment in establishments associated with the same EIN and the number of establishments within a firm as the number of establishments that report using a common EIN. To classify a firm's industry, we assign to a firm the average self-reported, 6-digit NAICS code of its establishments so that the firm is classified in the same way as its establishments are on average.

A firm's entry date is measured as the date at which the QCEW records a non-zero number of workers associated with a particular EIN after four consecutive quarters of recording zero workers. A firm's exit date is measured as the last date at which the QCEW records a non-zero number of workers associated with a particular EIN prior to four consecutive quarters of recording zero workers. A firm's age is measured by tracking firms after entering. Upon entry, the firm is assigned an age of 1 quarter and the firm's age is incremented by 1 quarter for each period that it does not exit.

D.2 LEHD measures of size-rank relationships

D.2.1 Confidentiality protection for size-rank statistics in LEHD

Characterizing the employment distribution by firm rank, we used the Employer Characteristics File maintained by the U.S. Census Bureau's Longitudinal Employer-Household Dynamics Program. A number of steps were taken to minimize the disclosure risk associated with the release of statistics on the upper ranks of the firm size distribution. First, instead of a direct size-rank regression, we coarsened the underlying distribution, employing finer categories the closer we are at the firm size distribution. Coarsening ensures that there are a large number of observations in each cell, even at the upper range of the distribution in which we use the finest categories. To limit disclosure risk further, we estimated a fifth-order polynomial (plus a constant) on the (average) percentile rank associated with each rank category. To conform to U.S. Census Bureau disclosure requirements, all point estimates and standard errors were rounded to four decimal points.

Table D.2.1: Fifth-order polynomial approximations of size-rank relationships

	Employment			Establishments			Establishment Size		
	1995	2005	2014	1995	2005	2014	1995	2005	2014
Intercept	16.33	9.586	4.932	11.55	14.11	17.37	13.87	7.406	2.725
	(0.04863)	(0.04889)	(0.04988)	(0.06931)	(0.06404)	(0.06247)	(0.04857)	(0.04882)	(0.04979)
$ln(rank)$	-6.478	0.4619	5.243	-6.302	-8.406	-11.36	-5.143	1.333	6.056
	(0.04041)	(0.4060)	(0.04145)	(0.05760)	(0.5319)	(0.05191)	(0.04036)	(0.04055)	(0.04137)
$ln(rank)^2$	2.224	-0.2067	-1.888	2.274	2.961	3.919	1.723	-0.5921	-2.268
	(0.01241)	(0.1247)	(0.01273)	(0.01769)	(0.01633)	(0.01594)	(0.01239)	(0.01245)	(0.01270)
$ln(rank)^3 \times 10$	-3.936	-0.09348	2.583	-4.121	-5.157	-6.550	-3.055	0.6863	3.381
	(0.01785)	(0.01793)	(0.0183)	(0.02544)	(0.02348)	(0.02292)	(0.01782)	(0.01790)	(0.01827)
$ln(rank)^4 \times 10^2$	3.206	0.3817	-1.605	3.354	4.078	5.012	2.528	0.2700	-2.283
	(0.01219)	(0.01225)	(0.01250)	(0.0174)	(0.01604)	(0.0157)	(0.01218)	(0.01223)	(0.01248)
$ln(rank)^5 \times 10^3$	-9.863	-0.2010	0.3571	-0.9933	-1.184	-1.419	0.7974	0.009459	0.5591
	(0.003199)	(0.003213)	(0.003280)	(0.00456)	(0.004209)	(0.004108)	(0.003195)	(0.003208)	(0.003274)

Notes: Authors' calculations of LEHD microdata. Standard errors are in parentheses. Estimates characterize March of each respective year. All point estimates and standard errors were rounded to four decimal places to conform to U.S. Census Bureau discloure requirements. The dependent variables employment, establishments, and average establishment size are in logs.

D.2.2 Polynomial estimation procedure and results

We now describe the polynomial approximations that allow us to characterize the distributions of three measures of firm size: total employment, the number of establishments, and average establishment size. We first ranked firm-level data by each of these different size measures. Using these ranks, we started with the smallest rankings, and assigned categories based an observation being within percentile ranges. The ranges are defined as follows.

1. Starting with the lowest, group observations into 1% bins until the 95th percentile is reached, for a total of 95 categories.
2. Group observations into 0.5% bins until the 99th percentile is reached, for a total of 8 categories.
3. Group observations into 0.1% bins until the 99.9th percentile is reached, for a total of 9 categories.
4. Group the remaining observations into 0.01% bins, for a total of 10 categories.

Using this method creates a total of 122 categories. This method of grouping the data was meant to provide a balance between generating information that can be informative about the tails of the distributions that we are interested in, which protecting the confidentiality of the underlying microdata: even the finest cells have a relatively large number of observations (e.g., 0.01% $\times$ 5 million = 500 observations). Each bin was assigned its average percentile rank (e.g., the lowest bin has an average percentile rank of 0.5, the next has an average percentile rank of 1.5, etc.). Polynomial approximations of our size measures use a transformation of this: $log((100 - \text{average percentile}) \times 1000)$. The transformation times 1000 was done for computational reasons, but conceptually is just a simple shift of the intercept because $log((100 - \text{average percentile}) \times 1000) = log(100 - \text{average percentile}) + log(1000)$. Fifth-order polynomials of this transformation of the average percentile rank (plus a constant) serve as regressors for each size measure.

One set of dependent variables consist of the lograrithm of each size measure for size in March of 1995, 2005, and 2014. To avoid approximating the discrete jumps in the distribution between small values (1, 2, 3, etc.) a random draw form the interval $[-0.50.5]$ is applied to each observation. Results of these regressions are shown in Appendix Table D.2.1. A second set of dependent variables show the nature of firm entry and growth in the year 2005 and employ a polynomial approximation of $100 -$ average percentile. Specifically, growth is defined as the change in the log of the size measure between March of 2004 and March of 2005. Entry is measured as positive employment in March of 2005 and zero employment in March of 2004. Results of this second set of regressions is shown in Appendix Table D.2.2.

D.2.3 Recovering the size distribution from the polynomial approximations

While our polynomial approximations capture much of the rich features of the underlying microdata, they are difficult to immediately interpret. We therefore transform these polynomial estimates to

Table D.2.2: Fifth-order polynomial approximations of rank relationships, 2005

	Growth Rates			
	Employment	Establishments	Establishment Size	Entry
$Intercept \times 10^2$	6.614	4.218	2.396	1.035
	(0.1477)	(0.1363)	(0.1973)	(0.1067)
$rank \times 10^3$	6.065	-5.522	11.59	10.91
	(0.3037)	(0.2802)	(0.4056)	(0.2157)
$rank^2 \times 10^4$	-4.786	0.2644	-7.430	6.573
	(0.1906)	(0.1758)	(0.2545)	(0.1337)
$rank^3 \times 10^5$	0.1430	0.5658	1.996	1.797
	(0.004878)	(0.04501)	(0.06514)	(0.0339)
$rank^4 \times 10^7$	1.884	0.5498	2.434	2.041
	(0.04517)	(0.04998)	(0.07235)	(0.03738)
$rank^5 \times 10^9$	0.8485	0.1983	1.470	0.8275
	(0.02169)	(0.02001)	(0.02896)	(0.01488)

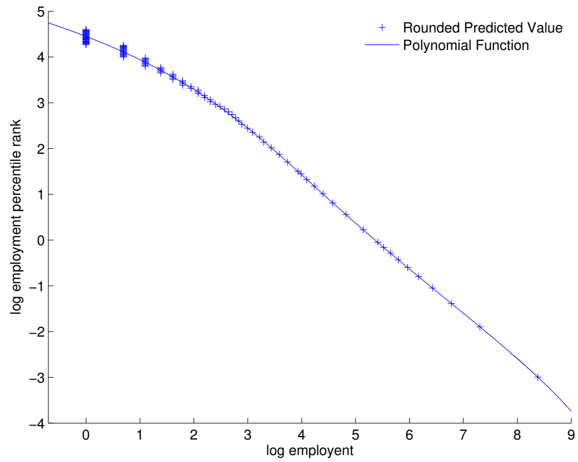
Notes: Authors' calculations of LEHD microdata. Estimates compare March of 2004 to March of 2005. Entry implies positive employment in 2005 but not 2004. Growth compares 2005 employment to 2004 employment for those with positive employment in both years. All point estimates and standard errors were rounded to four decimal places to conform to U.S. Census Bureau disclosure requirements.

create tables and figures that highlight important features of the underlying microdata. One feature of the data that we wish to highlight is that our estimation was done on coarsened (discritized) data. There are maximum and minimum values, and the are discrete values that the independent variable takes. Another feature of the underlying microdata for our size distribution is that total employment and the number of establishments are discrete, while our polynomials are of course continuous. While in practice average establishment size takes values other than integers, at most points in the distribution there are almost no multi-establishment firms and so the average establishment size is approximately a step function, especially for low values. To highlight these features of the data, we round each size measure after an exponential transformation.

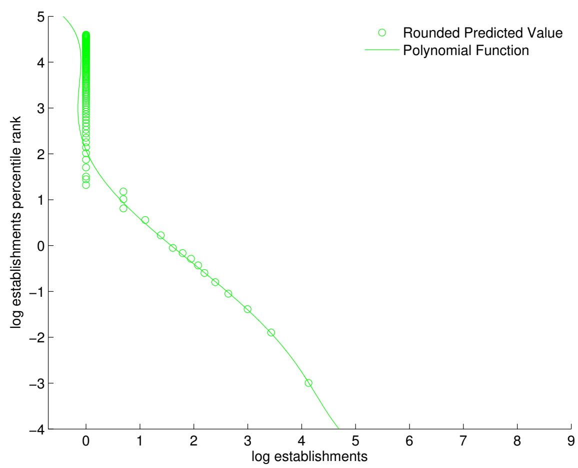
We show the relationship between the polynomials and the rounded predicted values in Appendix Figure D.2.1. This shows the relationship between the size measure and its ranked, and so each size measure is ranked separately. Percentile ranks, on the vertical axis, are given according to $log(100 - \text{average percentile})$. The log of the size measure is on the horizontal axis. The polynomial function extends beyond the predicted values. Unrounded values, by construction, all lie on the polynomial function. The rounded fitted values are at the upper end of the distribution almost identical with the polynomial. At the lowest values of log employment, the rounded values are more visible, especially for low values such as $log(1) = 0$, $log(2) \approx 0.69$, $log(3) \approx 1.10$. Rounding the size measure produces the feature that a disproportionate number of businesses have a very low number of employees and establishments. This bunching at the low end of the distribution is especially apparent in Panel D.2.1(b), which shows the size-rank relationship for the number of establishments. About 96% of firms have exactly one (rounded) establishment, which in logs is zero. However the polynomial ranges from -0.2 ($e^{-0.2} \approx 0.8$) to 0.4 ($e^{0.4} \approx 1.49$, as higher values

Figure D.2.1: Polynomial Approximations and Rounded Predicted Values, 2005

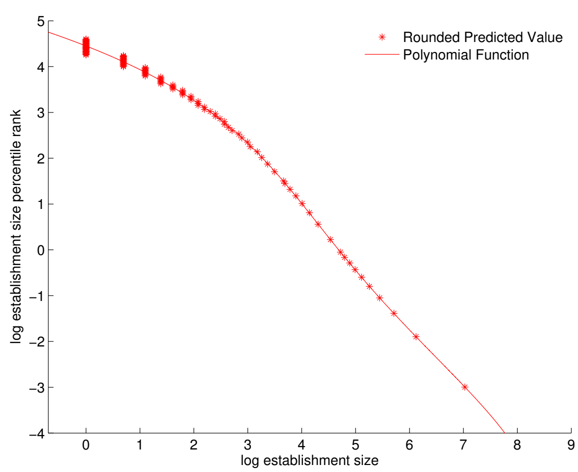
(a) Employment



(b) Establishments



(c) Establishment Size



Notes: Author's calculations of Longitudinal Employer-Household Dynamics microdata. Each series shows predicted values from a regression of the log size measure on a fifth order polynomial of the log percentile rank of the size measure in March of 2005, rounded to the nearest integer.

round to 2). The next three percentile bins round to two establishments, and thereafter each bin contains a distinct number of establishments. The total employment, shown in Panel D.2.1(a), as well as average establishment size, shown in Panel D.2.1(c) exhibit less dramatic bunching at small discrete values.

We also consider the relationships between employment rank and the number of establishments and establishment size. In this case, the size measure is not related to its rank, but to the rank of total employment. This is useful because when the data are so ranked, $log(\text{employment}) = log(\text{establishments}) + log(\text{establishment size})$. However, in this case, rounding no longer captures the salient features of moderate levels of employment (around employment of 10) where the number of establishments becomes distinct from zero. To capture these relationships, we fix total employment as the rounded value of total employment, and we estimate the number of establishments and average establishment size using Kuhn-Tucker optimization.

Let the log of the unrounded predicted value of total employment be e, the log number of establishments be p (plants) and the log average size be w (workers per establishment), both of which will be estimated based on unrounded predicted values $\hat{p}$ and $\hat{w}$, respectively. Specifically, we minimize the squared distance between the estimated value and the polynomial approximation subject to three constraints. First, total log employment is the sum of the logs of the number of establishments and workers per establishment, and so $e = p + w$, and the value of this constraint is μ. Second, the total log number of plants must be at least one and so $p \geq 0$, which has value λ_1. Third, the total log number of workers per plant must be at least one and so $w \geq 0$, which has value λ_2. The Kuhn-Tucker problem is now:

$$\min_{p,w,\mu,\lambda_1,\lambda_2} (p - \hat{p})^2 + (w - \hat{w})^2 + \mu(e - p - w) + \lambda_1 p + \lambda_2 w$$

subject to the inequality constraints $p \geq 0$ and $e \geq 0$.

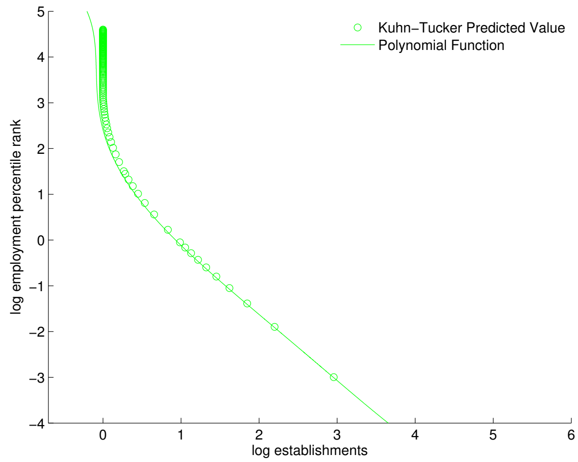
The full set of conditions that characterize the solution to this problem are:

$$\begin{aligned}
2(p - \hat{p}) - \mu - \lambda_1 &= 0 \\
2(w - \hat{w}) - \mu - \lambda_2 &= 0 \\
e - p - w &= 0 \\
\lambda_1 p &= 0 \\
\lambda_2 w &= 0 \\
p &\geq 0 \\
w &\geq 0.
\end{aligned}$$

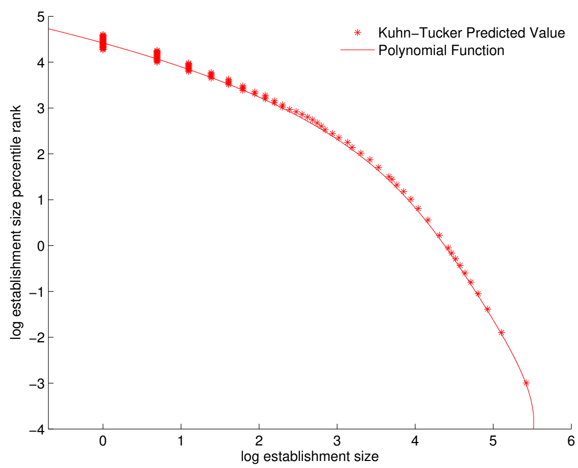
At an interior solution, $p > 0$ and $w > 0$, and so $\lambda_1 = \lambda_2 = 0$. In this case, $2(p - \hat{p}) = 2(w - \hat{w})$

Figure D.2.2: Polynomial Approximations and Kuhn-Tucker Optimization, 2005

(a) Establishments



(b) Establishment Size



Notes: Author's calculations of Longitudinal Employer-Household Dynamics microdata. Each series shows predicted values from a regression of the log size measure on a fifth order polynomial of the log percentile rank of employment in March of 2005, adjusted via Kuhn-Tucker optimization.

and we can substitute the constraint $e = p + w$ to recover

$$p = \frac{e + \hat{p} - \hat{w}}{2}$$
$$w = \frac{e - \hat{p} + \hat{w}}{2}.$$

Otherwise, at least one of the inequality constraints is binding and the solution is set to minimize the criterion function. In practice, this means that the number of establishments is set to zero when log total employment is greater than zero and an interior solution does not hold. The rounded employment counts e^* are then used to generate the final estimated number of establishments p^*

and workers per establishment w^* according to:

$$p^* = e^* \times \frac{p}{p+w}$$
$$w^* = e^* \times \frac{w}{p+w}.$$

The results of this optimization procedure are shown in Appendix Figure D.2.2. The estimated values from the Kuhn-Tucker optimization procedure for the number of establishments p^* and workers per establishment w^* are shown, as well as the analogous predicted values from the polynomial approximation $\hat{p}$ and $\hat{w}$, and the polynomial estimate itself. The inequality constraint binds on 77% of the distribution, or until total employment is about 10. At this point in the total employment distribution, a tiny fraction of firms are estimated to have multiple establishments as an interior solution and the inequality constraint no longer binds. Therefore, the number of establishments grows more continuously in Appendix Panel D.2.2(a) than the rounded establishment counts in Appendix Panel D.2.1(b). The effects of rounded employment can be seen in the discrete jumps at very low values of employment, especially those corresponding to total employment of 1 or 2.

D.2.4 Estimation of the average slope of the size-rank relationship

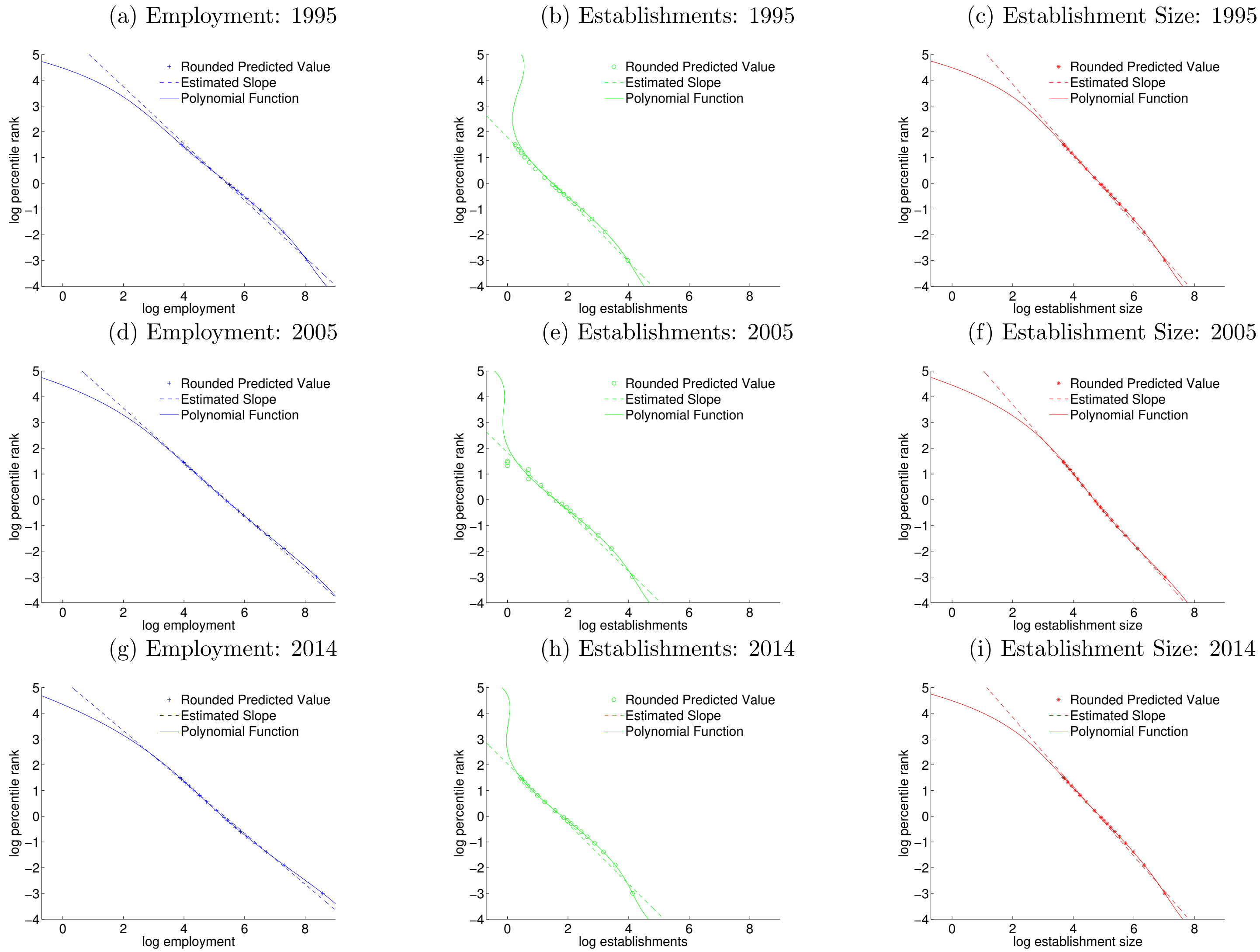
In order to characterize the slope of the tail of the size-rank relationship, we regress $log(100 -$ average percentile) on the log of each of our size measures. We use the 27 fitted values for the categories corresponding to the 95th percentile and above of the size distribution. Results of this estimation procedure are shown in Appendix Table D.2.3. We also compare the slopes of these estimates and the fitted values in Appendix Figures D.2.3 and D.2.4.

Table D.2.3: Size-rank regression estimates

	Employment	Establishments	Establishment Size
		95th percentile and above	
		Estimates for 1995	
Intercept	5.96	1.79	6.54
	(0.11)	(0.05)	(0.09)
Slope	-1.10	-1.20	-1.35
	(0.02)	(0.02)	(0.02)
		Estimates for 2005	
Intercept	5.66	1.83	6.40
	(0.12)	(0.06)	(0.10)
Slope	-1.05	-1.15	-1.35
	(0.02)	(0.02)	(0.02)
		Estimates for 2014	
Intercept	5.31	2.04	6.17
	(0.19)	(0.10)	(0.22)
Slope	-0.99	-1.17	-1.32
	(0.03)	(0.03)	(0.04)
		99th percentile and above	
		Estimates for 1995	
Intercept	6.54	1.95	6.85
	(0.18)	(0.10)	(0.17)
Slope	-1.17	-1.25	-1.39
	(0.02)	(0.03)	(0.03)
		Estimates for 2005	
Intercept	5.88	2.01	6.27
	(0.24)	(0.12)	(0.20)
Slope	-1.08	-1.20	-1.33
	(0.03)	(0.03)	(0.03)
		Estimates for 2014	
Intercept	5.27	2.21	5.66
	(0.39)	(0.22)	(0.40)
Slope	-0.99	-1.21	-1.24
	(0.05)	(0.06)	(0.06)

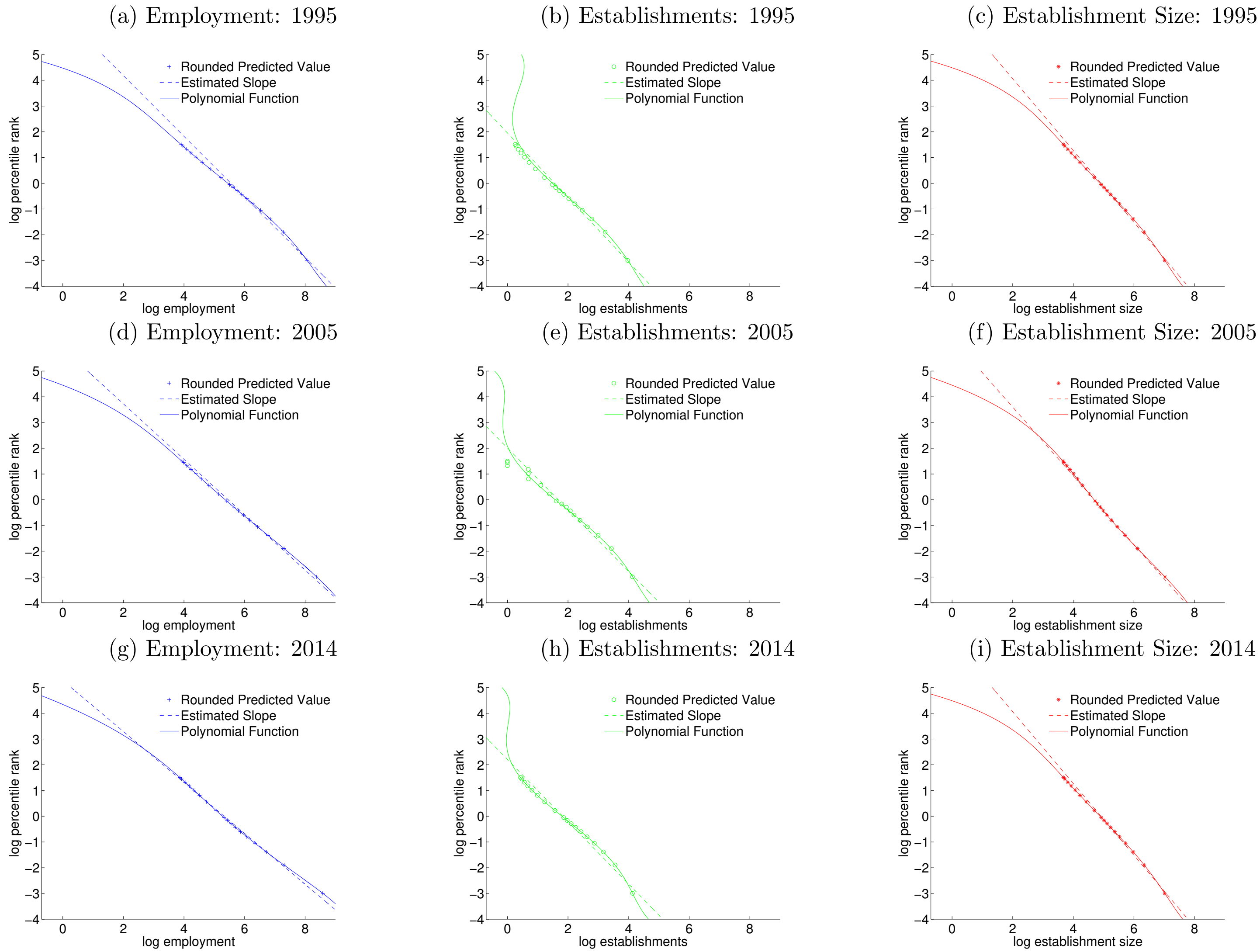
Notes: Authors' calculations of LEHD microdata. Linear regression of log outcome on log rank for fitted values that are at or above the 95th and 99th percentiles, respectively. Estimates reference March of each year.

Figure D.2.3: Size-Rank Regression Slope of 95th Percentile and Above



Notes: Author's calculations of Longitudinal Employer-Household Dynamics microdata. Each series shows predicted values from a regression of the log size measure on a fifth order polynomial of the log percentile rank of the size measure in March of 2005.

Figure D.2.4: Size-Rank Regression Slope of 99th Percentile and Above



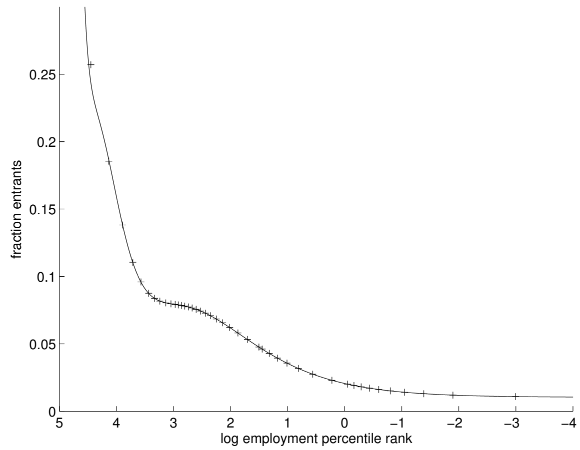
Notes: Author's calculations of Longitudinal Employer-Household Dynamics microdata. Each series shows predicted values from a regression of the log size measure on a fifth order polynomial of the log percentile rank of the size measure in March of 2005.

Figure D.2.5 shows the polynomial approximation for the share of entrants by log percentile rank, along with the fitted values from that polynomial approximation, for 2005. The estimated tails appear smooth at the extremes of the distribution. The fitted values for the fraction entrants take the average of all observations that round to discrete employment counts: 1 employee, 2 employees, etc. This binds at the lower tail of the distribution. For example, 25.7% of firms with exactly one employee are entrants, and these have an average percentile rank of 14.5. This is illustrated in Figure D.2.5 with value $log(100 - 14.5) \approx 4.45$. This aggregated value indicates a higher fraction entrants than the entry rate of 24.4% reported for log percentile rank 4.45. Entry rates for bins that round to higher discrete values of employment are closer between the predicted values and the polynomial approximation. Establishments with exactly two employees have an average rank of 37.5, illustrated with $log(100 - 37.5) = 4.14$, and a predicted value of 18.7, which is close to the polynomial value of 18.6. Larger firms exhibit almost no difference in the predicted entry rates between aggregated categories and the polynomial itself.

Published aggregates from the Bureau of Labor Statistics (BLS) provide an opportunity to benchmark the number of establishments distribution for 2005.[42] Figure D.2.6 compare the approximated values from the LEHD estimation with the BLS published aggregates on a log-size, log-percentile rank scale. The BLS published aggregates indicate that 95.2% of firms have only one establishment. This implies that the upper 4.8% of the number of establishments distribution has two or more establishments. We illustrate this in our log-log plot with the point $(log(2) \approx 0.69, log(4.8) \approx 1.57)$. The LEHD data suggest that a smaller share of firms, 3.7%, have multiple establishments. The BLS published aggregates indicate that 2.5% of firms have three or more establishments, while the LEHD data indicate a share of 2.0%. The BLS published aggregates indicate that 1.8% of firms have four or more establishments, while the LEHD data indicate a share of 1.4%. The BLS published aggregates indicate that 1.4% of firms have four or more establishments, while the LEHD data indicate a share of 1.1%. The BLS aggregates also indicate that 0.6% of firms have ten or more establishments, while the LEHD data indicate a share of 0.5%. Fitting a line to these five data points provide somewhat different slopes for the log-size, log-rank relationship. The natural explanation for these differences is that our LEHD microdata is for a 28-state subset, while the BLS published aggregates are national. Despite a level difference, the slopes are similar. The slope of this relationship is -1.27, while that of the LEHD using these same five data points is -1.23. The steeper slope of the BLS published aggregates suggests a somewhat thinner tail. Yet both of these estimated slopes are close to the slope estimates of -1.15 and -1.21 for the 95th, and 99th percentiles and above, respectively. These results, reported in Table D.2.3, include many more percentiles for the upper percentiles of the distribution, and so we naturally prefer these when we target moments for estimation.

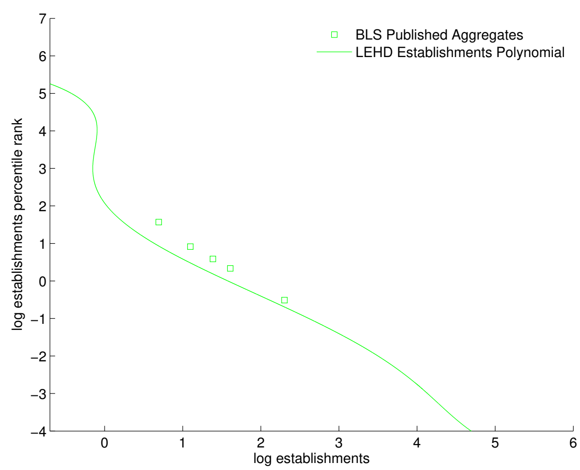
[42] Numbers for 2005 Q1 are taken from `https://www.bls.gov/bdm/sizeclassqanda.htm` (last accessed: October 4, 2019).

Figure D.2.5: Fraction entrants: polynomial approximation and predicted values, 2005



Notes: Author's calculations of Longitudinal Employer-Household Dynamics microdata. Each series shows predicted values from a regression of the log size measure on a fifth order polynomial of the log percentile rank of the size measure in March of 2005, rounded to the nearest integer.

Figure D.2.6: Comparison of Polynomial Approximation to Published Establishment Number Totals for 2005



Notes: Author's calculations of Longitudinal Employer-Household Dynamics microdata and Bureau of Labor Statistics published aggregates.

D.3 Estimation of parametrized establishment size distributions

In this Appendix, we describe an estimation strategy for recovering model parameters from establishment size distributions. We assume that the data is drawn from simple parametric distributions that are known to fit the actual U.S. data from past studies. We first estimate these distributions using publicly available data on establishment size distributions. This procedure provides us the data moments from establishment size distributions that we will substitute by the moments that directly come from our dataset once the disclosure process is complete. The data moments include the Pareto tail index of the establishment size distribution which cannot be inferred directly from publicly available data. The tail index is crucial for our estimation procedure.

We assume that in year t, the distribution of establishment size in number of employees (call it $\mathbf{l}$) takes the following form:

$$\Pr_t(\mathbf{l} \leq l) = G(\log l; \mu_t^e, \sigma_t^e, \lambda_t^e),$$

for establishment sizes $\ell = 1, 2, \ldots$ and in years $t = 1995, 2014$, where G is the CDF of the convolution between a normal distribution and an exponential distribution (see Sager and Timoshenko (2019) for more details on this type of distribution):

$$G(z; \mu, \sigma, \lambda) \equiv \Phi_n\left(\frac{z-\mu}{\sigma}\right) - e^{-\lambda(z-\mu)+\frac{\sigma^2}{2}\lambda^2}\Phi_n\left(\frac{z-\mu-\lambda\sigma^2}{\sigma}\right),$$

where Φ_n is the cdf of the standard normal distribution. This distribution flexibly nests both a normal distribution and an exponential distribution and conveniently allows for a thick right tail.

We estimate μ_t^e, σ_t^e, and λ_t^e by targeting Table D.3.1 published by BLS (second and forth column) for 1995 and 2014 using weighted least square minimization procedure in Sager and Timoshenko (2019). The estimation yields

$$\mu_{1995}^e = 0.6642, \quad \sigma_{1995}^e = 1.3488, \quad \lambda_{1995}^e = 1.4028$$

and

$$\mu_{2014}^e = 0.7170, \quad \sigma_{2014}^e = 1.3896, \quad \lambda_{2014}^e = 1.4423.$$

The targeted moments are closely replicated as shown in Table D.3.1 published by BLS (third and fifth column) and Figure 7. The estimated parameters are similar to those using Census' establishment data in Kondo et al. (2019). Notice that the establishment size distribution has a tail index that increases over time ($\lambda_{2014}^e > \lambda_{1995}^e$), indicating that skewness in the establishment size distribution has declined over time.

Table D.3.1: Distribution of establishments by employment

Establishment Size Bin	% of total, 2014 Data	Synthetic	% of total, 1995 Data	Synthetic
(a) 1-4	48.94	50.35	49.15	51.36
(b) 5-9	21.11	19.98	22.62	20.16
(c) 10-19	14.26	13.89	13.87	13.65
(d) 20-49	9.80	9.98	8.95	9.50
(e) 50-99	3.27	3.34	3.00	3.09
(f) 100-249	1.86	1.73	1.71	1.57
(g) 250-499	0.47	0.44	0.43	0.40
(h) 500-999	0.18	0.16	0.16	0.15
(i) 1000+	0.10	0.10	0.09	0.09
Mean size	17.51	18.04	16.87	17.31

E Additional empirical facts

E.1 Cross-sectional employment patterns

To further explore how firm size is related to intensive and extensive margins, in Figure E.1 we plot the average of establishment size at each firm (extensive margin) and the average of the number of establishments at each firm (intensive margin) in different firm size bins.[43] When all size measures are ranked by employment, this ensures that log employment is equal to the sum of the log number of establishments and the log average establishment size. The number of establishments distribution has a much steeper slope when ranked this way, which indicates that the largest firms do not always have the most establishments. Yet the number of establishments plays a very important role at the upper end of the distribution where its size-rank relationship grows rapidly.

Denoting the firm size by $\mathbf{Z}$, the number of establishments per firm as $\mathbf{X}$, and the average establishment size for each firm as Y, one can decompose the firm size into extensive and intensive margin:

$$\log(\mathbf{Z}) = \log(\mathbb{E}[\mathbf{X}|\mathbf{XY} = \mathbf{Z}]) + \log(\mathbb{E}[\mathbf{Y}|\mathbf{XY} = \mathbf{Z}]) + \Omega, \tag{48}$$

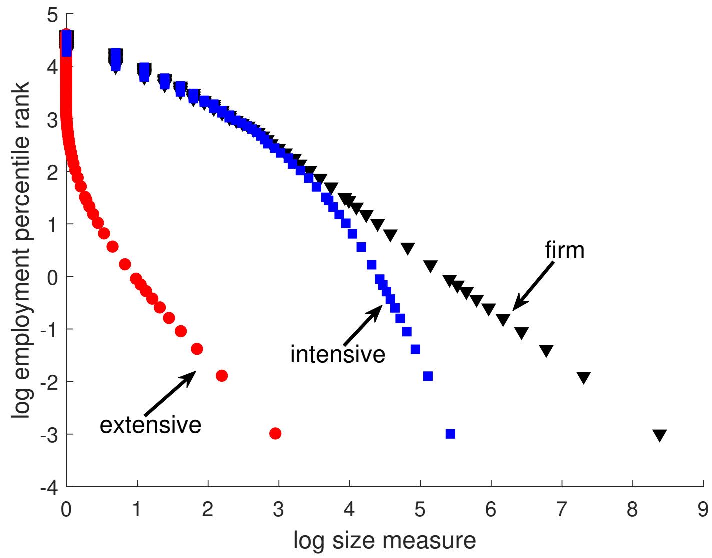
where $\Omega \leq 0$ and it is equal to zero if and only if $var[\mathbf{X}|\mathbf{XY} = \mathbf{Z}] = 0$.[44] In Figure 2, the left-hand side of (48) is the horizontal axis, and the first and the second terms are extensive and intensive margins. The sum of the slopes of the extensive and intensive margins can be different from one because of the Ω term (which varies with $\mathbf{Z}$).[45]

[43] Figure E.1 is constructed from polynomial approximations of the LEHD microdata. For the set of polynomial estimates from which these are constructed, see Appendix Table D.2.1.

[44] The derivation of (48) is in Appendix A.

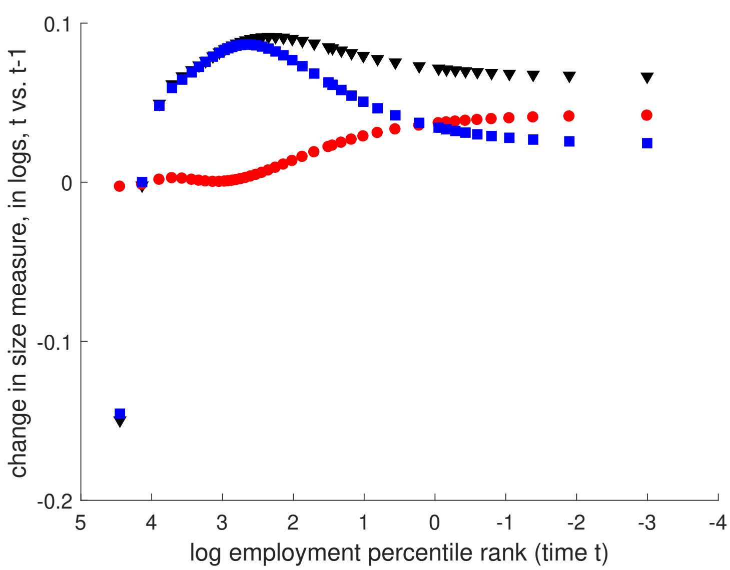
[45] Xi (2016) draws a similar figure to Figure 2, although his graph describes different pattern for intensive margin averages. He is not explicit about how his graph is drawn (including the data source), but we obtain a similar graph to his once we put the second and the third term in (48) together instead of our intensive margin concept.

Figure E.1: Size-rank relationships, ranked by employment



Notes: Author's calculations of Longitudinal Employer-Household Dynamics microdata. Each series shows fitted values from a regression of the log size measure on a fifth order polynomial of the log percentile rank of employment in March of 2005. The employment series is rounded to the nearest integer. The number of establishments and average establishment size are adjusted via Kuhn-Tucker optimization. See text for additional details.

Figure E.2: Growth rates, ranked by employment



Notes: Author's calculations of Longitudinal Employer-Household Dynamics microdata. Each series shows predicted values from a regression of the log change from 2004-2005 on a fifth order polynomial of the log percentile rank of employment in March of 2005. Percentiles of the employment distribution that round to the same integer value are grouped together.

Derivation of Equation (48): First note that,

$$\begin{aligned}\log(\mathbf{Z}) &= \log(\mathbb{E}[\mathbf{XY}|\mathbf{XY}=\mathbf{Z}]) \\ &= \log(\mathbb{E}[\mathbf{X}|\mathbf{XY}=\mathbf{Z}]) + \log(\mathbb{E}[\mathbf{Y}|\mathbf{XY}=\mathbf{Z}]) + \log\left(\frac{\mathbb{E}[\mathbf{XY}|\mathbf{XY}=\mathbf{Z}]}{\mathbb{E}[\mathbf{X}|\mathbf{XY}=\mathbf{Z}]\mathbb{E}[\mathbf{Y}|\mathbf{XY}=\mathbf{Z}]}\right).\end{aligned}$$

Call the final term as Ω. Because $\mathbb{E}[\mathbf{XY}|\mathbf{XY}=\mathbf{Z}]=\mathbf{Z}$ and $\mathbb{E}[\mathbf{X}|\mathbf{XY}=\mathbf{Z}]\mathbb{E}[\mathbf{Y}|\mathbf{XY}=\mathbf{Z}] = \mathbb{E}[\mathbf{X}|\mathbf{XY}=\mathbf{Z}]\mathbb{E}[\mathbf{Z}/\mathbf{X}|\mathbf{XY}=\mathbf{Z}] = \mathbb{E}[\mathbf{X}|\mathbf{XY}=\mathbf{Z}]\mathbb{E}[1/\mathbf{X}|\mathbf{XY}=\mathbf{Z}]\mathbf{Z}$. From Jensen's inequality,

$$\mathbb{E}[\mathbf{X}|\mathbf{XY}=\mathbf{Z}]\mathbb{E}\left[\frac{1}{\mathbf{X}}\middle|\mathbf{XY}=\mathbf{Z}\right] \geq \mathbb{E}[\mathbf{X}|\mathbf{XY}=\mathbf{Z}]\frac{1}{\mathbb{E}[\mathbf{X}|\mathbf{XY}=\mathbf{Z}]} = 1,$$

and thus $\Omega \leq 0$ and the equality holds when $var[\mathbf{X}|\mathbf{XY}=\mathbf{Z}]=0$.

E.2 Firm Growth in the Cross-Section

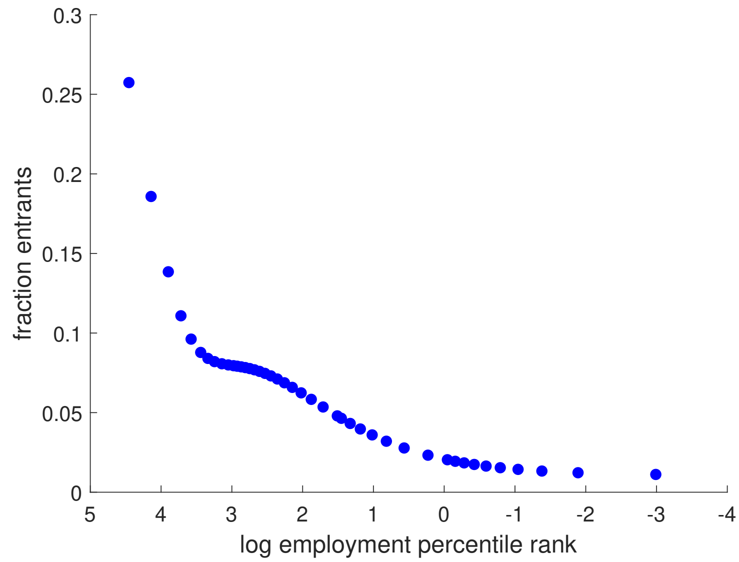
In Figure E.2 we document firm growth patterns in terms of extensive and intensive margins by comparing employment in March of 2005 to March in 2004, ranking firms by their size in 2005.[46] One-employee firms are likely to have contracted, and two-employee firms had zero growth. Incumbent firms with three or more workers are likely to have grown over the course of the year, by 5% to 9%. For most firms, growth is accomplished through expanding existing establishments rather than adding to the total. This is because nearly all (96%) of firms have exactly one establishment, and firms only occasionally transition between having a single establishment and multiple establishments. The largest firms grow through both expanding existing establishments as well as adding them. At around the upper percentile (i.e., a log percentile rank of zero) the number of establishments contribution dominates and more growth in total employment occurs from adding establishments than expanding them. Figure E.2 also shows that the smallest firms (with only one employee) often experienced a contraction. Note that -0.15 change in log employment corresponds to the average incumbent single-worker firm losing about a sixth of a worker in the previous year.

We measure the entry rates by firm type in Figure E.3.[47] For the smallest firms employing just one worker, about a third of them entered in the last year. Larger firms are less likely to have entered in the past year, and the probabilities for the upper percentile are close to zero and any positive probability can likely be attributed to measurement error.

[46] Figure E.2 is constructed from polynomial approximations of the underlying LEHD microdata. The polynomial estimates are contained in Appendix Table D.2.2.

[47] Figure E.3 is constructed from polynomial approximations of the underlying LEHD microdata. The polynomial estimates are contained in Appendix Table D.2.2.

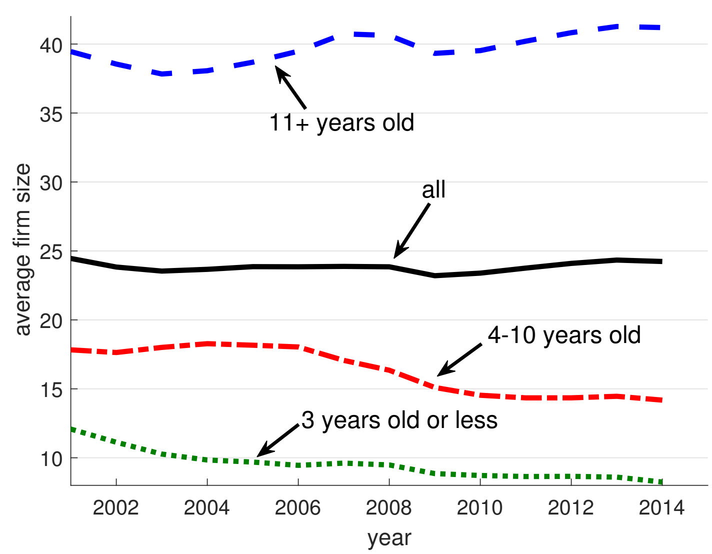
Figure E.3: Entry rates, ranked by employment



Notes: Author's calculations of Longitudinal Employer-Household Dynamics microdata. Each series shows predicted values from a regression of the entry rate on a fifth order polynomial of the log percentile rank of employment in March of 2005. Percentiles of the employment distribution that round to the same integer value are grouped together.

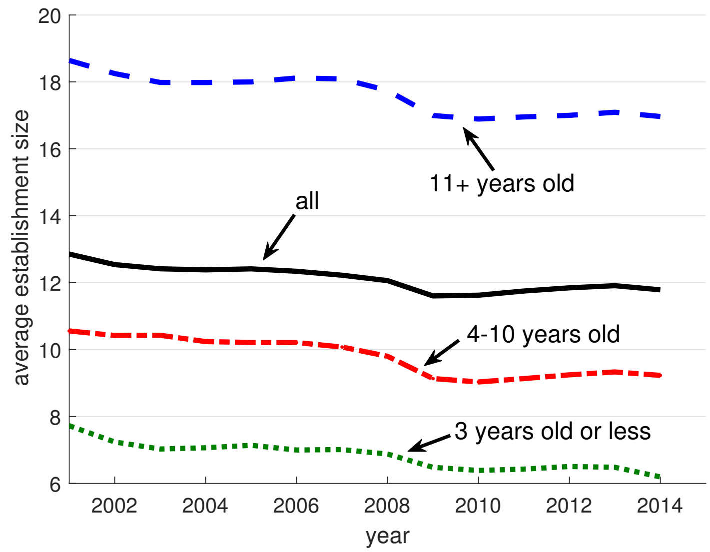
E.3 Additional time-series facts

Figure E.3.1: Average firm size (number of workers), different age groups



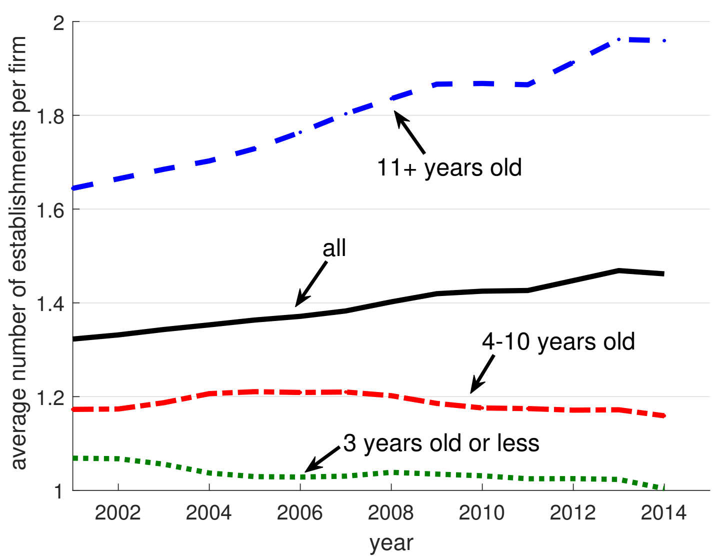
Notes: Author's calculations of Quarterly Census of Employment and Wages microdata.

Figure E.3.2: Average establishment size (number of workers), different age groups



Notes: Author's calculations of Quarterly Census of Employment and Wages microdata.

Figure E.3.3: Average establishment size (number of workers), different age groups



Notes: Author's calculations of Quarterly Census of Employment and Wages microdata.

We look at the outcomes for different ages of the firm. Note that the graphs for the age groups starts at the year 2001, so that we are able to consistently measure the age groups across different time periods. Figures E.3.1, E.3.2, and E.3.3 repeat the same plots as earlier. Because the significant increase in the firm size occurs during 1990s, Figure E.3.1 does not exhibit any obvious trend over the sample period for overall firms. However, we can see that there still is an increasing trend for the firms that are older than 11 years old. As in the case for all firms, the intensive margin does not exhibit an increasing trend in Figure E.3.2 for any age group. In contrast, Figure E.3.3 reveals a striking contrast across different age groups: the increase in extensive margin for overall economy is driven by the older firms. This motivates our modeling choice—firms are not born with different sizes across different time period; rather, their pattern of growth has changed over time.